The American Economic Review

ARTICLES

66

MARCH 1973

THE AMERICAN ECONOMIC ASSOCIATION

Founded in 1885

• Published at George Banta Co., Inc., Menasha, Wisconsin.

• The American Economic Review, including four quarterly numbers, the *Proceedings* of the annual meetings, and *Directory* and Supplements, is published by the American Economic Association and is sent to all members five times a year, in March, May, June, September, and December.

• Membership dues of the Association are $20.00 a year, which includes a year's subscription to both the *American Economic Review* and the *Journal of Economic Literature*. A three year junior membership of $10.00 per year is available to registered students. Subscriptions by nonmembers are $30.00 a year, and only subscriptions to both publications will be accepted. Single copies of the *Review* and *Journal* are $4.00 each. Each order for copies of either publication must also include a $.50 per order service charge. Orders should be sent to the Secretary's office, Nashville, Tennessee.

• Correspondence relating to the *Papers and Proceedings*, the *Directory*, advertising, permission to quote, business matters, subscriptions, membership and changes of address may be sent to the secretary, Rendigs Fels, 1313 21st Avenue, South, Nashville, Tennessee 37212. To be effective, notice of change of address must reach the secretary by the 1st of the month previous to the month of publication. The Association's publications are mailed by second class and are not forwardable by the Post Office.

• Second-class postage paid at Nashville, Tennessee and at additional mailing offices. Printed in U.S.A.

THE AMERICAN ECONOMIC REVIEW

March 1973

VOLUME LXIII, NUMBER 1

• Manuscripts and editorial correspondence relating to the regular quarterly issue of this REVIEW should be addressed to George H. Borts, Managing Editor of THE AMERICAN ECONOMIC REVIEW, Brown University, Providence, R.I. 02912. Manuscripts should be submitted in duplicate and in acceptable form and should be no longer than 50 pages of double-spaced typescript. A $10 submission fee must accompany each manuscript. *Style Instructions* for guidance in preparing manuscripts will be provided upon request to the editor.

• No responsibility for the views expressed by authors in this REVIEW is assumed by the editors or the publishers, The American Economic Association.

Articles

Shorter Papers

THE AMERICAN ECONOMIC REVIEW

JUNE 1973

VOLUME LXIII, NUMBER 3

• Manuscripts and editorial correspondence relating to the regular quarterly issue of this REVIEW should be addressed to George H. Borts, Managing Editor of THE AMERICAN ECONOMIC REVIEW, Brown University, Providence, R.I. 02912. Manuscripts should be submitted in duplicate and in acceptable form and should be no longer than 50 pages of double-spaced typescript. A $10 submission fee must accompany each manuscript. *Style Instructions* for guidance in preparing manuscripts will be provided upon request to the editor.

Articles

Shorter Papers

// THE AMERICAN ECONOMIC REVIEW

• Manuscripts and editorial correspondence relating to the regular quarterly issue of this REVIEW should be addressed to George H. Borts, Managing Editor of THE AMERICAN ECONOMIC REVIEW, Brown University, Providence, R.I. 02912. Manuscripts should be submitted in duplicate and in acceptable form and should be no longer than 50 pages of double-spaced typescript. A $10 submission fee must accompany each manuscript. *Style Instructions* for guidance in preparing manuscripts will be provided upon request to the editor.

• No responsibility for the views expressed by authors in this REVIEW is assumed by the editors or the publishers, The American Economic Association.

September 1973

VOLUME LXIII, NUMBER 4

Articles

Shorter Papers

THE AMERICAN ECONOMIC REVIEW

• Manuscripts and editorial correspondence relating to the regular quarterly issue of this REVIEW should be addressed to George H. Borts, Managing Editor of THE AMERICAN ECONOMIC REVIEW, Brown University, Providence, R.I. 02912. Manuscripts should be submitted in duplicate and in acceptable form and should be no longer than 50 pages of double-spaced typescript. A $10 submission fee must accompany each manuscript. *Style Instructions* for guidance in preparing manuscripts will be provided upon request to the editor.

• No responsibility for the views expressed by authors in this REVIEW is assumed by the editors or the publishers, The American Economic Association.

December 1973

VOLUME LXIII, NUMBER 5

Articles

Shorter Papers

THE AMERICAN ECONOMIC REVIEW

VOLUME LXIII

THE AMERICAN ECONOMIC ASSOCIATION
Executive Office: Nashville, Tennessee
Editorial Office: Brown University, Providence, Rhode Island

CONTENTS OF ARTICLES AND SHORTER PAPERS

CONTENTS OF PAPERS AND PROCEEDINGS

CONTRIBUTORS TO ARTICLES AND COMMUNICATIONS

CONTRIBUTORS TO PAPERS AND PROCEEDINGS

EMPLOYMENT SERVICES

NATIONAL REGISTRY FOR ECONOMISTS

The National Registry for Economists was established in January, 1966, to provide a centralized nationwide clearinghouse for economsts on a year-round basis. It is located in the Chicago Professional Placement Office of the Illinois State Employment Service and is staffed by experienced placement personnel, operating under the guidance and direction of Regional and National Bureau of Employment Security Professional Placement officials, and in cooperation with the American Economic Association. It is a free service. The National Registry for Economists maintains completely separate listings from those of the American Economic Association, and the National Registry does *not* have the listings as shown in the *American Economic Review*. There are no registration, referral, or placement fees. Application and order forms used in the Registry are available upon request from the: National Registry for Economists, Professional Placement Center, 40 West Adams Street, Chicago, Illinois 60603.

AMERICAN ECONOMIC ASSOCIATION
VACANCIES AND APPLICATIONS

The Association renders this service to economists who wish to make known their availability for positions and to those who are seeking to fill vacancies in the field of economics. The Association takes no responsibility for selecting applicants or for following up the results. The Secretary's office will merely afford a central point for forwarding inquiries. The *Review* will publish a description of vacancies and applications submitted, with necessary editorial changes. The Secretary would appreciate receiving notification of appointments made as a result of this service. All inquiries about a listing with a key number should refer to it specifically. Communications should be addressed to: The Secretary, American Economic Association, 809 Oxford House, 1313 21st Avenue South, Nashville, Tennessee 37212.

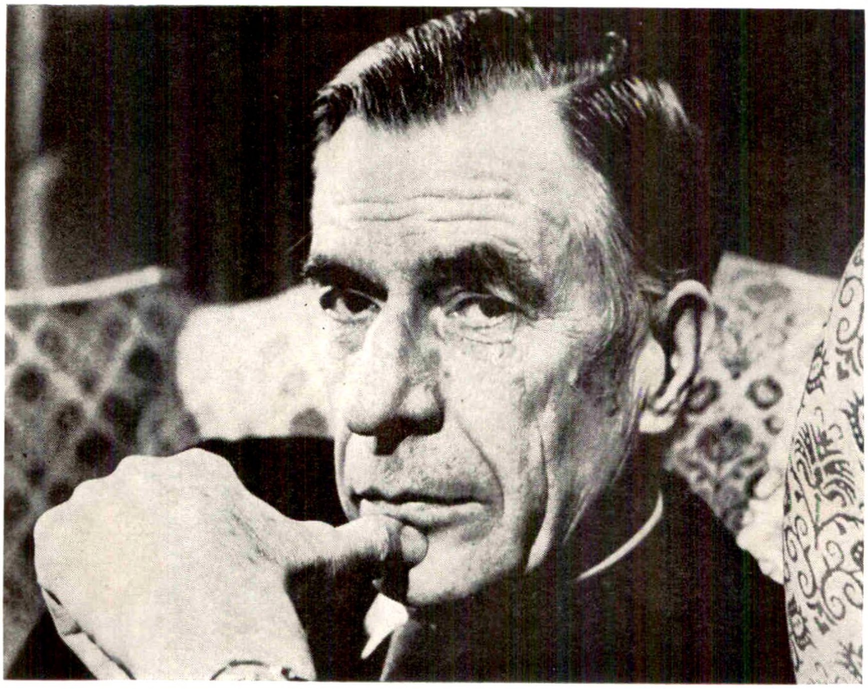
Résumés and application blanks are *not* supplied by the American Economic Association. The Secretary's office will merely forward any announcements of positions or résumés received to a designated key number. Deadlines for the four issues of the *Review* are January 1, April 1, July 1, and October 1. Announcements will normally be printed in two consecutive issues.

Vacancies

Academic Appointments: School of Comparative Social Sciences, Universiti Sains Malaysia (formerly known as the University of Penang); Senior Lecturers/Lecturers in international economics, regional economics, mathematical economics/econometrics/economic statistics, monetary and fiscal economics, development planning, intermediate economic theory, business administration. Salary range: Senior Lecturers from $22,380 (US $7932) to $30,120 (US $10680) per annum; Lecturers, from $16,800 (US $5964) to $26,580 (US $9420) per annum. Fringe benefits include return air passages and free medical coverage for appointee and family. Write to: The Registrar, Universiti Sains Malaysia, Minden, Pulau Pinang, Malaysia.

Economic Fellows: Senator Bill Brock announces fellowships for advanced graduate students, college professors or other professionals actually working within their area of expertise. Fellows would be expected to provide substantive materials leading to policy statements and legislation. Financial assistance is sometimes available for fellows at the rate of $4,000 to $5,000 a year. Fellows will generally remain in Senator Brock's office for at least six months. Application dates are open. Please send curriculum vitae, example of written work, and a description of your substantive interests to: Harrison Fox, Office of Senator Bill Brock, 304 Old Senate Office Building, Washington, D.C. 20510.

Number 74 of a series of photographs of past presidents of the Association



Power and the Useful Economist

By John Kenneth Galbraith*

The ceremonial address of the President of the American Economic Association is an art form which, I imagine like most of my predecessors, I have thoughtfully reviewed. On occasion, in the past, the addresses have dealt with some substantive problem of our subject or some afflicting problem of the economy. More often they have dealt, always a shade critically, with the methodology of economics. While accepting the larger science there has been adverse comment on the detailed elements of its practice. Economics is insufficiently normative. Model building has become an end, not a means. For several recent years in succession the criticism—which involved a certain element of personal introspection—included an exceptionally grave attack on mathematical economics. The style of these addresses, I might note in passing, is as distinctive as the subject matter. It features the thoughtful solemnity of men who sense that we are speaking for the ages. It may be worth a moment's time, on these great occasions, to recall that ours is a subject which features defeated expectations.

I am moved this evening to depart from the established rites. I should like to concern myself with basic questions of assumption and structure. If this breaks with tradition, it does not break with present professional tendency. We meet at a time when criticism is general—when the larger body of established theory is under extensive attack. Within the last half-dozen years what before was simply called economics in the nonsocialist world has come to be designated neoclassical economics with appropriate overtures to the Keynesian and post-Keynesian development. From being a general and accepted theory of economic behavior this has become a special and debatable interpretation of such behavior. For a new and notably articulate generation of economists a reference to neoclassical economics has become markedly pejorative.

I would judge as well as hope that the present attack will prove decisive. The established theory has reserves of strength. It sustains much minor refinement which does not raise the question of overall validity or usefulness. It survives strongly in the textbooks although even in this stronghold one senses anxiety among the more progressive or commercially sensitive authors. Perhaps there are limits to what the young will accept.

And the arrangements by which orthodoxy is conserved in the modern academy also remain formidable. In its first half century or so as a subject of instruction and research, economics was subject to censorship by outsiders. Businessmen and their political and ideological acolytes kept watch on departments of economics and reacted promptly to heresy, the latter being anything that seemed to threaten the sanctity of property, profits, a proper tariff policy, a balanced budget, or which involved sympathy for unions, public ownership, public regulation or, in any organized way, for the poor. The growing power and self-confidence of the educational estate, the formidable and growing complexity of our subject and, no doubt, the increasing acceptability of our ideas has largely relieved us of this intervention.

* Presidential address delivered at the eighty-fifth meeting of the American Economic Association, Toronto, Canada, December 29, 1972.

In leading centers of instruction faculty responsibility is either secure or increasingly so. But in place of the old censorship has come a new despotism. That consists in defining scientific excellence as whatever is closest in belief and method to the scholarly tendency of the people who are already there. This is a pervasive and oppressive thing not the less dangerous for being, in the frequent case, both self-righteous and unconscious.

But there are problems even with this control. Neoclassical or neo-Keynesian economics, though providing unlimited opportunity for demanding refinement, has a decisive flaw. It offers no useful handle for grasping the economic problems that now beset the modern society. And these problems are obtrusive—they will not lie down and die as a favor to our profession. No arrangement for the perpetuation of thought is secure if that thought does not make contact with the problems that it is presumed to solve.

I will not omit this evening to mention the failures of neoclassical theory. But I want also to urge the means by which we can reassociate ourselves with reality. Some of this will summarize past argument, more a book that is presently to be published. At this stage even the most conservative among my listeners will be reassured. To speak well of one's own published and unpublished writing, whatever one's other aberrations, is strongly in our professional tradition.

I

The most commonplace features of neoclassical and neo-Keynesian economics are the assumptions by which power, and therewith political content, is removed from the subject. The business firm is subordinate to the instruction of the market and, thereby, to the individual or household. The state is subordinate to the instruction of the citizen. There are exceptions but these are to the general and controlling rule, and it is firmly on the rule that neoclassical theory is positioned. If the business firm is subordinate to the market—if that is its master—then it does not have power to deploy in the economy save as this is in the service of the market and the consumer. And the winning of action to influence or rig the behavior of markets apart, it cannot bring power to bear on the state for there the citizen is in charge.

The decisive weakness in neoclassical and neo-Keynesian economics is not the error in the assumptions by which it elides the problem of power. The capacity for erroneous belief is very great, especially where it coincides with convenience. Rather in eliding power—in making economics a nonpolitical subject—neoclassical theory, by the same process, destroys its relation with the real world. The problems of this world, moreover, are increasing both in number and in the depth of their social affliction. In consequence neoclassical and neo-Keynesian economics is relegating its players to the social sidelines where they either call no plays or urge the wrong ones.

Specifically the exclusion of power and the resulting political content from economics causes it to foretell only two intrinsic and important economic problems. One of these is the microeconomic problem of market imperfection—more specifically of monopoly or oligopoly in product or factor markets—leading to aberration in resource and income distribution. The other is the macroeconomic problem of unemployment or inflation—of a deficiency or excess in the aggregate demand for goods and services, including that associated with monetary effects. And on both problems the failure is dramatic. Neoclassical economics leads to the wrong solution of the microeconomic problem and to no solution of the macroeconomic

problem. Meanwhile it leaves a whole galaxy of other urgent economic issues largely untouched.

It is now the considered sense of the community, even of economists when unhampered by professional doctrine, that the most prominent areas of market oligopoly—automobiles, rubber, chemicals, plastics, alcohol, tobacco, detergents, cosmetics, computers, bogus health remedies, space adventure—are areas not of low but of high development, not of inadequate but of excessive resource use. And there is a powerful instinct that in some areas of monopoly or oligopoly, most notably in the production of weapons and weapons systems, resource use is dangerously vast.

In further contradiction of the established microeconomic conclusions, we have an increasing reaction by the community to deficient resource use in industries that, at least in the scale and structure of the firm, approach the market model. Housing, health services, and local transportation are among the leading cases. The deprivation and social distress that follow from the poor performance of these industries are also something that, in their nondoctrinal manifestation, most economists take for granted.

The defender of the established doctrine does, of course, argue that excess and deprivation in resource use in the areas just mentioned reflect consumer choice. And in the areas of deprivation he can rightly insist that the fault lies with firms that, though small, are local monopolies or reflect the monopoly power of unions. These explanations beg two remarkably obvious questions: Why does the modern consumer increasingly tend to insanity, increasingly insist on self-abuse? And why do little monopolies perform badly and the big ones too well?

In fact the neoclassical model has no explanation of the most important microeconomic problem of our time. That problem is why we have a highly unequal development as between industries of great market power and industries of slight market power, with the development, in defiance of all doctrine, greatly favoring the first.[1]

The macroeconomic failure has been, if anything, more embarrassing. Save in its strictly mystical manifestation in one branch of monetary theory, modern macroeconomic policy depends for its validity and workability on the neoclassical market. That market, whether competitive, monopolistic, or oligopolistic, is the ultimate and authoritative instruction to the profit-maximizing firm. When output and employment are deficient, policy requires that aggregate demand be increased; this is an instruction to the market to which firms in turn respond. When the economy is at or near the effective capacity of the plant and the labor force and inflation is the relevant social discomfort, the remedy is reversed. Demand is curtailed; the result is either an initial effect on prices or a delayed one as surplus labor seeks employment, interest rates fall and lower factor costs bring stable or lower prices.

Such is the accepted basis of policy. It follows faithfully from the neoclassical faith in the market. The practical consequences of pursuing it need no elucidation. It has been tried in recent years in every developed country. The common result has been politically unacceptable unemployment, persistent and (in my view) socially damaging inflation or, more often, a combination of the two. The extreme failure has been, not surprisingly, in the most advanced industrial country,

[1] It will be observed that performance in agriculture, an industry in which the firm has little market power, is not adverse. But it should also be observed that in no industry has power over prices been more completely removed to public authority or is there greater effort at collective control of costs or a more comprehensive socialization of technology.

the United States. But the recent experience of Britain has been almost equally disenchanting. One gathers that there may be Canadian politicians who now believe that a combination of unemployment and inflation is not the best platform on which to fight a general election.

We should not deny ourselves either the instruction or the amusement that comes from the recent history of the United States in this matter. Four years ago Mr. Nixon came to office with a firm commitment to neoclassical orthodoxy. In this he was supported by some of its most distinguished and devout exponents in all the land. His subsequent discovery that he was a Keynesian involved no precipitate or radical departure from this faith. The discovery came thirty-five years after *The General Theory*; as I have just noted, all neo-Keynesian policy rests firmly on the paramount role of the market. But then a year and a half ago, facing reelection, he found that his economists' commitment to neoclassical and Keynesian orthodoxy, however admirable in the abstract, was a luxury that he could no longer afford. He apostatized to wage and price control; so, with exemplary flexibility of mind, did his economists although admittedly this acceptance of the real world has still to survive its critical test which is the apostates' return to computers and classrooms. But our admiration for this pliability should not keep us from recalling that, when the President changed course, no American economists were anywhere working on the policy he was forced by circumstances to adopt. And it is even more disturbing that few are now working on the policy which we have been forced to follow.

More economists, in fact, are still concerning themselves with the effort to reconcile controls with the neoclassical market. This has involved an unrewarding combination of economics and archeology with wishful thinking. It holds that an inflationary momentum developed during the late 1960's in connection with the financing—or underfinancing—of the Vietnam war. And inflationary expectation became part of business and trade union calculation. The momentum and expectation still survive. The controls are necessary until these are dissipated. Then the neoclassical and neo-Keynesian world will return, along with the appropriate policies, in all their quiet comfort. We may be sure that will not happen. Nor will we expect it to happen if we see the role of power and political decision in modern economic behavior.

II

In place of the market system, we must now assume that for approximately half of all economic output there is a power or planning system. (The latter term seems to me more descriptive, less pejorative and thus preferable.) The planning system consists in the United States of, at the most, 2,000 large corporations. In their operation they have power that transcends the market. They rival where they do not borrow from the power of the state. My views on these matters will be familiar at least to some, and I shall spare myself the pleasure of extensive repetition. I cannot think that the power of the modern corporation, the purposes for which it is used or the associated power of the modern union would seem implausible or even very novel were they not in conflict with the vested doctrine.

Thus we agree that the modern corporation either by itself or in conjunction with others, has extensive influence over its prices and its major costs. Can we doubt that it goes beyond its prices and the market to persuade its customers? Or that it goes back of its costs to organize supply? Or that from its earnings or the possession of financial affiliates it seeks to

control its sources of capital? Or that its persuasion of the consumer joined with the similar effort of other firms—and with the more than incidental blessing of neoclassical pedagogy—helps establish the values of the community, notably the association between well-being and the progressively increased consumption of the products of this part of the economy?

And as citizens, if not as scholars, we would not deny that the modern corporation has a compelling position in the modern state. What it needs in research and development, technically qualified people, public works, emergency financial support, becomes public policy. So does the military procurement that sustains the demand for numerous of its products. So, perhaps, does the foreign policy that justifies the military procurement. And the means by which this power is brought to bear on the state is widely accepted. It requires an organization to deal with an organization. And between public and private bureaucracies—between GM and the Department of Transportation, General Dynamics and the Pentagon—there is a deeply symbiotic relationship. Each of these organizations can do much for the other. There is even, between them, a large and continuous interchange of executive personnel.

Finally over this exercise of power and much enhancing it is the rich gloss of reputability. The men who guide the modern corporation, including the financial, legal, technical, advertising, and other sacerdotal authorities in corporate function, are the most respectable, affluent, and prestigious members of the national community. They are the Establishment. Their interest tends to become the public interest. It is an interest that even some economists find it comfortable and rewarding to avow.

That interest, needless to say, is profoundly concerned with power—with winning acceptance by others of the collective or corporate purpose. It does not disavow profits. These are important for ensuring the autonomy of the management—what I have called the technostructure—and for bringing the supply of capital within the control of the firm. Profits are also a source of prestige and therewith of influence. But of paramount importance is the much more directly political goal of growth. Such growth carries a strong economic reward; it directly enhances the pay, perquisites, and opportunities for promotion of the members of the technostructure. And it consolidates and enhances authority. It does this for the individual—for the man who now heads a larger organization or a larger part of an organization than before. And it increases the influence of the corporation as a whole.

Neoclassical economics is not without an instinct for survival. It rightly sees the unmanaged sovereignty of the consumer, the ultimate sovereignty of the citizen and the maximization of profits and resulting subordination of the firm to the market as the three legs of a tripod on which it stands. These are what exclude the role of power in the system. All three propositions tax the capacity for belief. That the modern consumer is the object of a massive management effort by the producer is not readily denied. The methods of such management, by their nature, are embarrassingly visible. It can only be argued that somehow it all cancels out. Elections in the United States and Canada are now being fought on the issue of the subordination of the state to corporate interest. As voters, economists accept the validity of the issue. Only their teaching denies it. But the commitment of the modern corporate bureaucracy to its expansion is, perhaps, the clearest of all. That the modern conglomerate always pursues profit over aggrandizement is believed by none. It is a commonplace of these last years, strongly

reflected in securities' prices, that agglomeration has always been good for growth but often bad for earnings.

There remains in the modern economy—and this I stress—a world of small firms where the instruction of the market is still paramount, where costs are given, where the state is remote and subject through the legislature to the traditional pressures of economic interest groups and where profit maximization alone is consistent with survival. We should not think of this as the classically competitive part of the system—in contrast with the monopolistic or oligopolistic sector from which the planning system has evolved. Rather, in its combination of competitive and monopolistic structures, it approaches the entire neoclassical model. We have, to repeat, two systems. In one, power is still, as ever, contained by the market. In another and still evolving system, power extends incompletely but comprehensively to markets, to the people who patronize them, to the state and thus, ultimately, to resource use. The coexistence of these two systems becomes, in turn, a major clue to economic performance.

III

Power being so comprehensively deployed in a very large part of the total economy, there can no longer, except for reasons of game-playing or more deliberate intellectual evasion, be any separation by economists between economics and politics. When the modern corporation acquires power over markets, power in the community, power over the state, power over belief, it is a political instrument, different in form and degree but not in kind from the state itself. To hold otherwise—to deny the political character of the modern corporation—is not merely to avoid the reality. It is to disguise the reality. The victims of that disguise are those we instruct in error. The beneficiaries are the institutions whose power we so disguise. Let there be no question: Economics, so long as it is thus taught, becomes, however unconsciously, a part of an arrangement by which the citizen or student is kept from seeing how he is, or will be, governed.

This does not mean that economics now becomes a branch of political science. That is a prospect by which we would rightly be repelled. Political science is also the captive of its stereotypes—including that of citizen control of the state. Also while economics cherishes thought, at least in principle, political science regularly accords reverence to the man who knows only what has been done before. Economics does not become a part of political science. But politics does—and must—become a part of economics.

There will be fear that once we abandon present theory, with its intellectually demanding refinement and its increasing instinct for measurement, we shall lose the filter by which scholars are separated from charlatans and windbags. These latter are always a danger, but there is more danger in remaining with a world that is not real. And we shall be surprised, I think, at the new clarity and intellectual consistency with which we see our world, once power is made a part of our system. To such a view let me now turn.

IV

In the neoclassical view of the economy a general identity of interest between the goals of the business firm and those of the community could be assumed. The firm was subject to the instruction of the community, either through the market or the ballot box. People could not be fundamentally in conflict with themselves—always given some reasonable decency in income distribution. Once the firm in the planning system is seen to have comprehensive power to pursue its own interest,

this assumption becomes untenable. Perhaps by accident its interests are those of the public but there is no organic reason why this must be so. In the absence of proof to the contrary, divergence of interest, not identity of interest, must be assumed.

The nature of the conflict also becomes predictable. Growth being a principal goal of the planning system it will be great where power is great. And in the market sector of the economy, growth will, at least by comparison, be deficient. This will not, as neoclassical doctrine holds, be because people have an amiable tendency to misunderstand their needs. It will be because the system is so constructed as to serve badly their needs and then to win greater or less acquiescence in the result. That the present system should lead to an excessive output of automobiles, an improbable effort to cover the economically developed sections of the planet with asphalt, a lunar preoccupation with moon exploration, a fantastically expensive and potentially suicidal investment in missiles, submarines, bombers, and aircraft carriers, is as one would expect. These are the industries with power to command resources for growth. And central to public purpose—to sound resource utilization—will be a cutback in such industries, as all instinct now suggests. Thus does the introduction of power as a comprehensive aspect of our system correct present error. Let us not fail to note that these are exactly the industries in which an uncomplicated neoclassical view of monopoly and oligopoly and of profit maximization at the expense of ideal resource use would, of all things, suggest an expansion of output. How wrong are we allowed to be!

The counterpart of excessive resource use in the planning system where power is comprehensively deployed is a relatively deficient resource use where power is circumscribed. Such will be the case in the part of the economy where competition and entrepreneurial monopoly as distinct from great organization are the rule. And if the product or service is closely related to comfort or survival, the discontent will be considerable. That housing, health services, local transportation, some household services, are now areas of grave inadequacy is agreed. It is in such industries that all modern governments seek to expand resource use. Here, in desperation, even the devout free enterprisers accept the need for social action, even of socialism.

Again, we may observe, the error of economics is prejudicial. Although as citizens we advocate restraint in the area of excessive resource use, our teaching does not. And though as citizens we urge social action where the firm approaches the neoclassical norm, our teaching does not. In this latter case we not only disguise corporate power but we make remedial action in such areas as housing, health care, transportation, also abnormal—the consequence of *sui generis* error that is never quite explained. This is unfortunate for here are tasks that require imagination, pride and determination.

V

When power is admitted to our calculus, our macroeconomic embarrassment also disappears. Economics makes plausible what governments are forced, in practice, to do. Corporations have power in their markets. So, and partly in consequence, do unions. The competitive claims of unions can most conveniently be resolved by passing the cost of settlement along to the public. Measures to arrest this exercise of power by limiting aggregate demand must be severe. And, not surprisingly, the power of the planning system has been brought to bear to exclude those macroeconomic measures that have a primary effect on that system. Thus monetary policy is entirely

permissible; that is at least partly because its primary effect is on the neoclassical entrepreneur who must borrow money. Monetary constraint is far less painful for the large established corporation which, as an elementary exercise of power, has ensured itself a supply of capital from earnings or financial affiliates or morally affiliated banks. The power of the planning system in the community has also won immunity for public expenditures important to itself—highways, industrial research, rescue loans, national defense. These have the sanction of a higher public purpose. A similar if still slightly less successful effort is being made on behalf of corporate and personal taxes. So fiscal policy has also been accommodated to the interests of the planning system.

Thus the inevitability of controls. The interaction of corporate and trade union power can be made to yield only to the strongest fiscal and monetary restraints. Those restraints that are available have a comparatively benign effect on those with power, but they weigh adversely on people who vote. When no election is in prospect, perhaps such a policy is possible. It will earn applause for its respectability. But it cannot be tolerated by anyone who must weigh its popular effect.

As with the need for social action and organization in the market sector there are many reasons why it would be well were economists to accept the inevitability of wage and price control. It would help keep politicians, responding to the resonance of their own past instruction, from supposing controls to be wicked and unnatural and hence temporary and to be abandoned whenever they seemed to be working. This is a poor mood in which to develop sound administration. And it would cause economists themselves to consider how controls can be made workable and how the effect on income distribution can be made equitable. With controls this last becomes a serious matter. The market is no longer a disguise for inequality, however egregious, in income distribution. Much inequality must be seen to be the result of relative power.

VI

When power is made part of our system, yet other matters of considerable current moment are illuminated. Thus the counterpart of systemic differences in development as between the planning and market sectors of the economy is systemic sectoral differences in income. In the neoclassical system, resource mobility is assumed, broadly speaking, to equalize inter-industry return. If there is inequality, it is the result of barriers to movement. Now we see that, given its comprehensive market power, the planning system can protect itself from adverse movements in its terms of trade. The same power allows it to accept unions for it need not absorb even temporarily their demands. In the market system, some areas of monopoly or union power apart, there is no similar control of the terms of trade. Given the absence of market power there can be no similar yielding on wage costs for there is no similar certainty that they can be passed on. (It is because of the character of the industry he seeks to organize, not his original power, that Cesar Chavez is for so many the new Lenin.) And, in the market system, the self-employed have the option —not present in the planning system—of reducing their own wages (and sometimes those of families or immediate employees) in order to survive.

Thus there is a built-in inequality in income between the two systems. And thus also the case for minimum wage legislation, support to trade unions in agriculture, price support legislation, and most important, perhaps, a floor under family income as antidotes to such inter-

industry inequality. Again this view of matters fits our present concerns. Minimum wage legislation, price support legislation, and support to collective bargaining are all questions of continuing political controversy as they apply to small business and agriculture. They are not serious issues in highly organized industry—in the planning system. And the question of a floor under family income, a matter of intense political argument, has recently divided workers in the planning system who would not be beneficiaries from those in the market system who would be. Again there is reassurance in a view of the economy that prepares us for the politics of our time.

The inclusion of power in economic calculus also prepares us for the great debate over the environment. It is the claim of neoclassical economics that it foresaw possible environmental consequences from economic development—that it, some time ago, embraced the concept of external diseconomies of production and, by inference, of consumption. Alas, this is a modest claim. The noninclusion of external diseconomies was long viewed as a minor defect of the price system—an afterthought for an hour's classroom discussion. And, as E. J. Mishan has observed, it was largely ignored in the textbooks. Nor does the notion of external diseconomies now offer a useful remedy. No one can suppose, or really supposes, that more than a fraction of the damage—especially that to the beauty and tranquility of our surroundings—could be compensated in any useful way by internalizing external diseconomies.

If growth is the central and rewarding purpose of the firm and if power is comprehensively available to impose this goal on the society, the possibility of conflict between private growth and public purpose as regards the environment is immediately plausible. So, since this power depends extensively not on force but persuasion, is the effort to make pollution seem palatable or worth the cost, including the effort to make advertising of remedial action a substitute for action. And so is the remedy to which all industrial countries are being forced. This is not, primarily, to internalize external diseconomies. Rather it is to specify the legal parameters within which growth may proceed or, as in the case of automobile use in central cities, airplane use over urban areas, the SST, industrial, commercial, and residential appropriation of countryside and roadside, the kinds of growth that are inconsistent with public purpose. We would have saved much corruption of our surroundings if our economics had held such action to be the predictable consequence of the pursuit of present economic goals and not the exceptional result of a peculiar aberration of the price system.

We had best, in any case, have the right guide to action for the future for there is a strong conservative case for such guidance. While economists toy weakly with external diseconomies, others are arguing that growth itself is the villain. They are seeking its extinction. To see environmental damage as a natural consequence of planning power and purpose and to see, in consequence, the need for confining growth within parameters that protect the public interest could be important for ensuring continued economic growth.

Finally, when power becomes part of our system, so does Ralph Nader. We are prepared for the explosion of concern now called consumerism. If the consumer is the ultimate source of authority, his abuse is an occasional fault. He cannot be fundamentally at odds with an economic system that he commands. But if the producing firm has comprehensive power and purposes of its own, there is every likelihood of conflict. Technology is then subordinate to the strategy of consumer persuasion.

Products are changed not to make them better but to take advantage of the belief that what is new is better. There is a high failure rate in engineering not what is better but what can be sold. The consumer—the unpersuaded or disenchanted consumer—rebels. This is not a rebellion against minor matters of fraud or misinformation. It is a major reaction against a whole deployment of power by which the consumer is made the instrument of purposes that are not his own.

VII

There are two conclusions to which this exercise—to which incorporation of power into our system—compels us. The first, in a way, is encouraging. It is that economists' work is not yet done. On the contrary, it is just beginning. If we accept the reality of power as part of our system, we have years of useful work ahead of us. And since we will be in touch with real issues, and since issues that are real inspire passion, our life will, again, be pleasantly contentious, perhaps even usefully dangerous.

The other conclusion concerns the state. For when we make power and therewith politics a part of our system, we can no longer escape or disguise the contradictory character of the modern state. The state is the prime object of economic power. It is captured. Yet on all the matters I have mentioned—the restrictions on excessive resource use, organization to offset inadequate resource use, controls, action to correct systemic inequality, protection of the environment, protection of the consumer—remedial action lies with the state. The fox is powerful in the management of the coop. To this management the chickens must look for redress.

Thus perhaps our greatest question. Is emancipation of the state from the control of the planning system possible? No one knows. And in the absence of knowledge no one certainly will suggest that it will be easy. But there is a gleam of encouragement. As ever circumstances are forcing the pace.

In the United States the recent election was fought, all but exclusively, over issues in which the purposes of the planning system or its major participants diverge from those of the public. The question of defense expenditures is such an issue. That of tax reform is another. The deprivation in housing, mass transportation, health services, city services, is yet another—one that reflects the relative inability of these industries to organize and command resources. The question of a guaranteed income is another such issue. Its effect, as I have noted, is on incomes outside the planning system—on the exploited in the market system, those who are rejected by both. The environment is such an issue—with its conflict between the technostructure's goal of growth and the public concern for its surroundings. Only wage and price control was not an issue in the recent election. That was almost certainly because economists of orthodox tendency on both sides found the prospect too embarrassing to discuss.

I do not mention these issues with any purpose save to show that the questions that emerge when power is made a part of our calculus are present and real. We need hardly remind ourselves that political issues are made not by parties and politicians but by circumstance.

Once power is made part of our system, we will not of course escape the political contention that comes from dealing with issues that are real. This brings me to my last point. I do not plead for partisanship in our economics but for neutrality. But let us be clear as to what is neutral. If the state must be emancipated from economic interest, a neutral economics would not deny the need. This is what economics now

does. It tells the young and susceptible and the old and vulnerable that economic life has no content of power and politics because the firm is safely subordinate to the market and to the state and for this reason it is safely at the command of the consumer and citizen. Such an economics is not neutral. It is the influential and invaluable ally of those whose exercise of power depends on an acquiescent public. If the state is the executive committee of the great corporation and the planning system, it is partly because neoclassical economics is its instrument for neutralizing suspicion that this is so. I have spoken of the emancipation of the state from economic interest. For the economist there can be no doubt as to where this task begins. It is with the emancipation of economic belief.

Short-Run Dynamics in Models of Money and Growth

By DOUGLAS D. PURVIS*

Originating with James Tobin's initial treatment of money as a second asset in the Solow one-sector growth model, the subject of money and growth has received a great deal of attention in the recent literature. Tobin's emphasis on portfolio balance to determine the equilibrium of the model provides a useful framework for the discussion of the development of two opposing "schools of thought" among recent writers on the subject of money and growth. The neoclassical approach follows Tobin in his emphasis on portfolio balance and includes contributions by Miguel Sidrauski and Harry Johnson, among others. A cogent and comprehensive statement of the neoclassical viewpoint can be found in the excellent survey by David Levhari and Don Patinkin. The second approach—the Keynes-Wicksell approach —as expounded by Jerome Stein in particular, and also including contributions by Hugh Rose and Keizo Nagatani, faults the neoclassical model on two basic and related points. They are the implications of the model for the dynamics of price change, and the lack of independent savings and investment decisions.

The neoclassical approach is characterized by the assumption that desired per capita real balances are always held—prices must adjust instantaneously to assure portfolio balance. A given rate of expansion of the nominal money stock combined with the exogenously given rate of population growth serves to determine the equilibrium rate of inflation consistent with asset equilibrium. The division of assets between money and physical capital is thereby determined, and there need be no specification of an independent investment function—all physical savings are instantaneously channelled into capital accumulation, and the desired capital stock is always held.[1]

The long-run properties of the neoclassical model allow for the coexistence of nonzero steady-state inflation and goods market equilibrium by specifying that excess demand for goods causes a departure from the steady-state rate of inflation, but is not a necessary condition for a nonzero inflation rate at any point in time. It is contended that, in a dynamic world, there are two forces operating to drive the price level—excess demand for goods and inflationary expectations. In steady state, excess demand is zero; actual inflation equals expected inflation, not necessarily zero; and the possibility of nonzero steady-

* Queen's University. This paper was written while I was a graduate student at the University of Chicago. My understanding of the issues has been greatly improved by many discussions with Rudiger Dornbusch, Stanley Fischer, and Michael Mussa. I also wish to thank members of workshops at the University of Chicago, University of Rochester, and York University, and to participants in the Chicago Symposium on Trade, Growth, and the Balance of Payments (University of Chicago, December 1970) for helpful comments on an earlier version of this paper. Jerome Stein, George Borts, and an anonymous referee provided very useful comments for which I am most grateful. Of course, I am responsible for any remaining errors. Financial support from the Canada Council is gratefully acknowledged.

[1] A puzzling result of Tobin's initial treatment is that the introduction of money into the barter model lowered the capital intensity and output per capita. I have recently tried to analyze this seemingly paradoxical result elsewhere (see Purvis).

state inflation represents a continuously changing equilibrium price level. However, the assumptions concerning portfolio balance outlined above impose rather implausible restrictions on the short-run behavior of the price level—prices must adjust instantaneously and completely to eliminate any disequilibrium. Hence when monetary disturbances are introduced into the analysis, the neoclassical model is seen to generate undesirable (or, at least, unrealistic) short-run movements in the price level.[2]

Accordingly, the essential features incorporated into the Keynes-Wicksell models are the specification of independent savings and investment decisions, and the explicit representation of a price equation in which prices move only in response to excess demand in the goods market. An immediate unfortunate consequence of such a model is that if monetary policy is such that the steady state is characterized by a nonzero rate of inflation, then there must be perpetual disequilibrium in the goods markets such that the plans of savers or investors, or both, must be frustrated continually. At this point the Keynes-Wicksell proponents need to resort to a *deus ex machina*—the institutionally determined allocation of frustrations—to close the model.

In a recent article, Stein has elucidated the differences between the two approaches, elaborating in particular on this very point of the different foundations of and implications for the dynamics of price change inherent in the two approaches. He apparently accepts the criticism outlined in the above paragraph, and outlines a proposal for reconciling the two approaches in which disequilibrium causes inflation to diverge from its expected value. Unfortunately, this analysis is hindered by a stock flow confusion arising from Stein's treatment of the Keynes-Wicksell model. There he (correctly) uses Walras' Law to relate price changes to the excess *flow* supply of money, but then chooses to treat the special case where a *flow* excess supply of money corresponds to a *stock* excess supply of money in such a manner that prices can be represented to move in response to a *stock* excess supply in the money market.[3] As we shall show below, this seemingly innocuous misspecification gives rise to some unnecessary complications in the analysis.

It is argued here that along with the specification of the determinants of inflation, the mere possibility of stock disequilibrium is in fact the crucial point of departure from the neoclassical scenario; and by explicitly considering the implications of stock disequilibrium on the demand behavior of the asset holders, the neoclassical and Keynes-Wicksell approaches are more easily reconcilable. Further, I hope to show that such considerations lead to the construction of a model which retains the desirable long-run properties of the neoclassical model, but which has more desirable short-run properties than either the neoclassical or Keynes-Wicksell models outlined above. This short-run behavior is derived from explicit analysis of flow equilibrium for assets in a model in which careful attention is given to the development of consistent stock-flow relationships under conditions of possible stock disequilibrium.

[2] That is, the neoclassical model represents a "frictionless" world in perpetual equilibrium. The present paper introduces adjustment costs whereby prices respond to eliminate asset disequilibrium only over time. On this, also see Fischer.

[3] Of course, the two are obviously closely related, but Stein's analysis implies the very restrictive case of a linear homogeneous relationship. Later in the present paper we incorporate a stock-adjustment term into the flow demand, and then Stein's case is only valid when the flow supply of real balances can be assumed constant. This seems most inappropriate for the analysis of short-run price behavior where shocks are induced through changes in the flow supply.

A simple neoclassical model is constructed in Section I, and its dynamics are briefly explained. In Section II adjustment costs are explicitly introduced into the neoclassical model in order to rationalize the possible existence of stock disequilibrium. Given such a disequilibrium, and given the assumed adjustment costs, wealth holders are seen to adjust their asset holdings along an optimal adjustment path which eliminates the stock disequilibrium *over time*. No uncertainty is involved—asset holders are assumed to be able to perceive the adjustment costs correctly and hence form flow demands that are feasible. Thus the adjustment process, or transitional stage, is characterized by flow equilibrium. It is important to note that while this transitional stage is also characterized throughout by stock disequilibrium, it is not correct to consider the analysis of the short-run behavior as "disequilibrium dynamics." It is equilibrium dynamics in the flow sense: given the initial stock disequilibrium, and given the assumed costs of adjusting these stocks, wealth holders adjust asset holdings in a *optimal* manner along an *equilibrium* path. Thus the analysis lends itself to the usual economic approach of considering a constrained optimization procedure giving rise to a set of equilibrium conditions.[4] Long-run equilibrium or steady state is restored when the stock disequilibrium is eliminated. Finally, in Section III, the special case analyzed by Stein is treated briefly, and some concluding comments are offered summarizing the implications of the reconciliation achieved.

I. A Simple Neoclassical Model

For simplicity, we will explore only the case where money is produced costlessly, held only by consumers, and distributed as a government transfer payment independently of existing real balances.[5] Per capita output is given by equation (1) and disposable income which includes changes in the value of existing real balances is given by equation (2):

$$(1)\qquad y = y(k)$$

$$(2)\qquad y_d = y(k) + m(\mu - \Pi)$$

where m is actual real balances per capita, k is the existing capital-labor ratio, μ is the given constant rate of increase in the nominal money stock, and Π is the actual rate of inflation. We assume that the labor force is growing at the constant exogenously given rate n, and that, for now, per capita consumption, c, is a constant fraction, $(1-s)$ of disposable income.[6]

Then, letting D denote the operator d/dt, and denoting physical savings per capita by S_p, we have:

$$(3)\qquad Dk = (y - c) - nk = S_p - nk$$

Portfolio equilibrium is achieved when the stock demand for real balances, given by equation (4), equals the stock supply.

$$(4)\qquad m^d = b(i)\cdot y(k);\quad b' < 0$$

In equation (4), desired per capita real balances are positively related to per capita output,[7] and the coefficient b

[4] The analogy to the stock-adjustment processes which have received a great deal of attention in the recent literature on the theory of investment of the firm is obvious. For example, see J. P. Gould.

[5] The latter assumption is necessary in order to retain the effects of monetary policy on the steady-state capital intensity. See Stein, p. 96.

[6] The limitations of this assumption are well known; however, it will suffice for our purposes except for the following important qualification. In a model incorporating stock disequilibrium, any semblance of maximizing behavior would indicate a savings rate varying with the stock disequilibrium. This is introduced in the savings function given by equation (13) in the text. Of course, a more satisfactory treatment would be a model of full dynamic optimization.

[7] It may be more correct to relate m^d to disposable income, but as the choice is inconsequential to my results, the present analysis is maintained for simplicity and for compliance with most of the literature.

varies inversely with the nominal interest rate i, defined as the sum of the real rate of interest $y'(k)$, plus the expected rate of inflation, Π^e.

$$i = y'(k) + \Pi^e$$

In order to concentrate later on the stock adjustment process, the simplifying but unrealistic assumption is made that Π^e always equals the equilibrium rate of inflation, $\Pi^e = \mu - n$.

Equilibrium, or steady state, occurs when $Dk = 0$, or, equivalently, when per capita physical savings S_p, given by equation (5), equals the investment requirement, nk.

$$(5) \quad \begin{aligned} S_p &= [s - (1 - s)\cdot b\cdot(\mu - \Pi)]y(k) \\ &= \sigma(k, \Pi)\cdot y(k) \end{aligned}$$

where $\sigma(k, \Pi)$ is the physical savings ratio; $\sigma_k < 0$, $\sigma_\Pi > 0$.

Following Sidrauski we set $S_p = nk$ and solve for equilibrium real balances, given by equation (6):

$$(6) \qquad m^* = \frac{sy(k^*) - nk^*}{(1 - s)n} = m(k^*)$$

where the asterisks denote equilibrium or steady-state values. Equation (6), plotted as the m^* curve in Figure 1, represents the (long-run) goods market equilibrium locus along which the per capita capital stock is constant. Above m^* there will be excess demand in the goods market and k will be falling, and conversely below m^*.

In Figure 1, $OR(\Pi_0)$ plots equation (4) for a given value of $\Pi^e = \Pi_0$; hence $OR(\Pi_0)$ is the locus of portfolio equilibrium along which $m(\equiv m^d)$ would be constant. Above and to the left of OR, m would be falling and conversely below OR. Of course, as emphasized above, the neoclassical model is characterized by perpetual asset equilibrium so these latter adjustments take place instantaneously via changes in the price level and all adjustment takes place

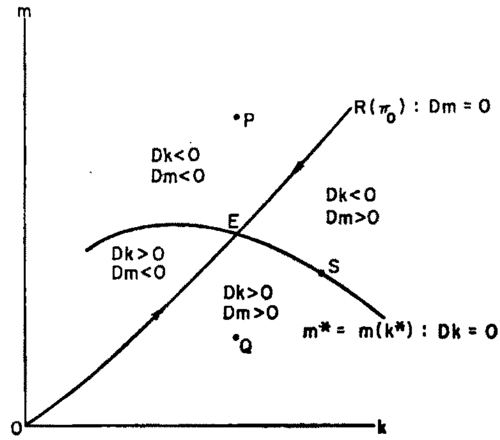


FIGURE 1

along $OR(\Pi_0)$ in Figure 1. Steady state, of course, attains at E.

A once-for-all increase in the nominal money supply would cause the economy to move momentarily to a position such as P, giving rise to an instantaneous proportionate increase in the price level causing it to return instantly to E. An increase in the rate of expansion of the nominal money stock would increase the equilibrium and expected rate of inflation causing OR to rotate clockwise; say now passing through points Q and S. Desired real balances would fall, hence prices would rise immediately to move the economy initially to a point such as Q. Then it would move along the new OR curve towards the new equilibrium position S.

The price dynamics of the neoclassical model can be seen by differentiating equation (4) with respect to time, and solving for Π, yielding:

$$(7) \qquad \Pi = \mu - n - \frac{B(k, \Pi^e)}{m}\cdot Dk$$

where $B(k, \Pi^e)$ is defined as $\partial m^d/\partial k$, i.e.,

$$B(k, \Pi^e) = [y(k)\cdot b'(i)\cdot y''(k) + b(i)\cdot y'(k)]$$

which is positive under usual assumptions on $y(k)$.

In the steady state, $Dk=0$ by definition, and hence equation (7) has the desirable steady-state solution $\Pi=\mu-n$. Further, from equation (7) it is found that in order for Π to exceed its steady-state value $\mu-n$, capital must be decumulating. That is, in terms of Figure 1, the section of *OR* northeast of *E* corresponds to a rate of inflation in excess of the steady-state rate, $\mu-n$ (i.e., real balances are falling), and conversely for the segment of *OR* southwest of *E*.

Before going on to consider the reconciliation of this model with the Keynes-Wicksell criticisms outlined above, it is perhaps worthwhile discussing briefly the attempt by Sidrauski to incorporate more reasonable short-run behavior into the neoclassical model by introducing adaptive expectations into the analysis. Since desired real balances are functionally dependent on the expected rate of inflation, gradual adjustment in expectations allows actual and desired real balances to remain equal; and at the same time, actual real balances can adjust more slowly towards their long-run equilibrium level following an exogenous monetary disturbance. Below the short-run behavior of our model is compared with that achieved by Sidrauski.

II. A Model With Stock Disequilibrium: A Proposed Neoclassical Variant

We wish now to relax the neoclassical assumption of perpetual stock equilibrium, in the sense that we no longer wish to treat m^d given by equation (4) as being identically equal to actual per capita real balances m. Such divergence between actual and desired real balances must ultimately derive from the existence of transactions costs in asset markets. In this section, it is assumed that there exist rising marginal costs to the rate of change of holdings of capital.[8] Then, as the wealth holder has only two asset forms in which he can hold wealth, and will choose to equalize own rates of return at each point in time, it is plausible to consider him at some point in time holding a stock of real balances different from his long-run desired stock. That is, we no longer constrain the price level to adjust instantaneously to eliminate any stock disequilibrium in the money market.

However, any stock disequilibrium will be eliminated over time as the wealth holder adjusts in an optimal manner. That is, the stock excess demand for real balances will affect the flow demand for real balances via some stock adjustment mechanism. As a first approximation, we postulate the following flow demand for real balances:

$$f^d = nm^d + B(k, \Pi^e) \cdot Dk - \alpha \cdot \bar{\bar{m}} \tag{8}$$

where the first term on the right-hand side of equation (8) is the steady-state flow demand, and the second term is the adjustment in desired real balances for changes in capital intensity as the economy moves towards the steady state.[9] The third term is the stock adjustment component of flow demand where $\bar{\bar{m}}$ is the stock excess supply of real balances $m-m^d$; and where α, the "speed of adjustment coefficient," has dimension time^{-1}. In order to simplify the

[8] This is obviously related to the Penrose effect in the theory of investment. See Hirofumi Uzawa and the references cited there. In the present case, the costs of changing wealth positions include, of course, consumption foregone.

[9] It seems unreasonable that wealth holders would demonstrate the kind of price expectations specified and still express demand functions in terms of current variables only. It would seem more appropriate to have them demonstrate similar degrees of sophistication in all markets. For example, the expectations function specified would be more consistent if demands were in terms of permanent incomes, while the present demand functions would be more consistent if combined with expectations $\Pi^*=\Pi$ (t). However, for present purposes we follow the case set out in the text. Although we refer to nm^d as the steady-state component, it differs from its steady-state value when the economy is out of steady state, i.e., when $m^d \neq m^*$. $B(k, \Pi^e)Dk$ represents the movement along the *OR* curve of Figure 1.

analysis while still emphasizing the stock-adjustment mechanism, we will treat α as a constant.[10]

The concept of the stock demand for real balances seems to warrant further comment at this point. The quantity m^d given by equation (4) defines, for given k and given Π^e, the stock that the asset holder would wish to be holding in the absence of any costs involved in adjusting his actual stock to that level, and m^d then can be regarded as the long-run desired stock of real balances. However, at any point in time there exists the distinct concept of the stock that the asset holder wishes to hold *at that instant*, given the existing stock (which may differ from the quantity m^d), and given the costs involved in changing the level of that existing stock. This latter concept can be regarded as the short-run (or instantaneous) desired stock of real balances, m^s.

While this distinction is useful in sorting out the concept of the stock demand for real balances, it is even more useful in clarifying the nature of the role played by the flow demand for real balances.[11] The long-run stock demand, m^d, except insofar as expectations are effected by asset flows, is defined independently of those flows. However, the short-run stock demand, m^s, and the flow demand are closely tied in the following way. By the very nature of the term $\alpha \cdot \bar{\bar{m}}$ in the flow demand given by equation (8), if real balances are growing at the desired rate then the asset holder is in short-run stock equilibrium. This is reconciled with the existence of stock disequilibrium in the long-run sense by the fact that the disequilibrium is being eliminated at the desired rate, and hence the wealth holder is, at that instant and in this limited sense, "happy" with his present stock of real balances. That is, m^s is the "effective" stock demand at any instant when the constraint imposed by the adjustment costs is taken into consideration; and the maintenance of flow equilibrium assures that m^s will identically equal actual m at every point in time. Then f^d will also equal the time derivative of $m^s (\equiv m)$.

Flow supply, f^s is given by equation (9), and the excess flow supply of real balances, $\bar{\bar{f}}$, by the difference $f^s - f^d$, in equation (10):

$$f^s = m(\mu - \Pi) \tag{9}$$

$$\bar{\bar{f}} = m(\mu - \Pi) - nm^d - B(k, \Pi^e) \cdot Dk + \alpha \bar{\bar{m}} \tag{10}$$

Now, although there may be stock disequilibrium at any point in time ($m^d \neq m$), wealth holders are able to add to their existing real balances at the desired rate—again, this is the nature of the term $\alpha \cdot \bar{\bar{m}}$ in equation (8). Then $\bar{\bar{f}}$ identically equals zero, and we can solve equation (10) for Π:

$$\Pi = \mu - n + \frac{\bar{\bar{m}}}{m} \cdot (n + \alpha) - \frac{B(k, \Pi^e)}{m} \cdot Dk \tag{11}$$

In the steady state, $\bar{\bar{m}} = Dk = 0$, so equation (11) has the desired steady-state solution, $\Pi = \mu - n$. For Π to exceed its steady-state value, $\bar{\bar{m}} \cdot (n + \alpha)$ must exceed $B(k, \Pi^e) \cdot Dk$. Now, contrary to the simple neoclassical model, it is possible for capital to be accumulating while the rate of inflation exceeds its steady-state value *if* there is a large enough excess stock supply of real balances.

If we define z as the excess of Π over its steady-state value, from equation (11) we

[10] A more complete analysis would undertake a full intertemporal dynamic optimization to determine the exact characteristics of the adjustment process—for an example of such an analysis in the theory of investment of the firm, see Gould.

[11] In terms of the language of the recent literature on the general theory of disequilibrium (see, for example, Herschell Grossman), m^d and m^s correspond closely to the concepts of "notional" and "effective" demand, respectively.

see that Dk is related to $\overline{\overline{m}}$ according to:

$$(12) \quad Dk = -\frac{1}{B(k, \Pi^e)} \cdot [zm - \overline{\overline{m}}(n + \alpha)]$$

Equation (12) is the flow equilibrium condition in the money market. For any given value of z, Dk is positively related to $\overline{\overline{m}}$ indicating the portfolio adjustment entailing dishoarding of real balances, and accumulation of physical capital. This relationship is plotted in Figure 2 where we get a family of positively sloped curves mm, each corresponding to different values of Π (or, equivalently, of z). For any point in $(Dk, \overline{\overline{m}})$ space there exists a value of Π consistent with flow equilibrium in the money market.

However, Dk must, in fact, move in accordance with equation (3)—the flow equilibrium condition in the goods market—except that we are no longer free to interpret that savings rate as a constant (compare fn. 6). To proceed, we need to explicitly introduce the long-run desired capital stock k^d. Of course, k^d cannot be specified independently of m^d since in the steady state when the long-run desired stocks are being held, the asset demands m^d and k^d must satisfy the wealth constraint, $w = m^d + k^d$. However, when we allow for stock disequilibrium, it is possible at any point in time to have an excess demand (or supply) for wealth in the sense that k^d plus m^d exceeds (falls short of) k plus m.[12] Then we wish to postulate the following behavior of the savings ratio:

$$(13) \qquad s = s(\overline{\overline{m}}, \overline{\overline{k}}); \quad s_1 < 0, s_2 < 0$$

where $\overline{\overline{k}}$ is the excess of actual k over k^d. The signs of the partial derivatives arise from the belief that an excess stock supply of either asset will induce people to consume a larger proportion of any given income, and hence the savings rate falls. We could have considered the simpler case where s depends only on the sum $\overline{\overline{m}} + \overline{\overline{k}}$, but the present analysis allows for the more general behavioral assumption that, although the two excess demands are obviously not independent, the asset holder may wish to adjust one at a different rate than the other. Of course, both versions yield the same steady-state solutions.

Then equation (3) can be rewritten as

$$(3') \qquad \begin{aligned} Dk &= \phi(\overline{\overline{m}}, \overline{\overline{k}}, \Pi); \\ \phi_1 &= s_1[y(k) + m(\mu - \Pi)] < 0 \\ \phi_2 &= s_2[y(k) + m(\mu - \Pi)] \\ &\quad + (sy' - n) \gtrless 0 \\ \phi_3 &= (1 - s) \cdot m > 0 \end{aligned}$$

Although the sign of ϕ_2 is theoretically ambiguous, it will be negative for all but very large negative values of $\overline{\overline{k}}$, and we shall treat ϕ_2 as negative throughout the analysis. Equation (3′), for $k^* = k$ and $\Pi = \Pi^e$ is plotted in Figure 2 as cc.[13] When we recognize that goods market equilibrium implies flow equilibrium in the money market, we can draw a goods market equilibrium curve for $k = k^*$ by allowing for the required change in Π. This is seen as $c'c'$ in Figure 2, which is flatter than cc due to the sign of ϕ_3. An increase

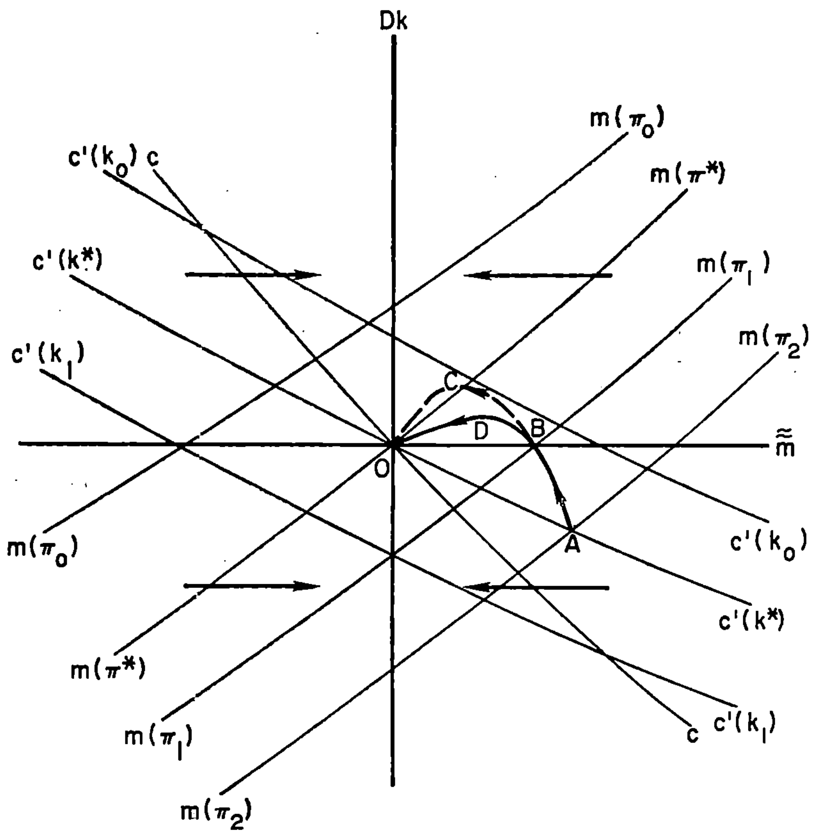
[12] In the text, w is interpreted as actual wealth, equals desired wealth in the steady state. Out of the steady state, it is $k^s + m^s$ that is constrained by actual wealth, $w = k + m$; where the interpretation of k^s is analogous to m^s. However, by the flow equilibrium postulate, $k^s - k = m^s - m = 0$, so that not only do the "short-run excess demands" sum to zero, but each individual short-run excess demand equals zero. Consideration of k^d demonstrates why the two Keynes-Wicksell criticisms of the neoclassical model cited in the second paragraph of the present paper are related. The neoclassical assumption of perpetual asset equilibrium extends, of course, to k^d; hence there is no need to also consider explicitly an investment function.

[13] The diagrammatic framework used here is not well suited for dynamic analysis—it would be more useful to work in state variable space. However, Figure 2 clearly sets out what is going on in the model, and how the portfolio adjustment and goods market equilibrium interact to determine Π and Dk.

in the capital stock would shift $c'c'$ down, and conversely for a fall in the capital intensity. Thus, again we get a family of curves, this time with negative slope each one plotting the equilibrium combinations of Dk and $\bar{\bar{m}}$ for a given capital intensity.

At any point in time m, p, k, μ, Π^e, and y are all given, and flow equilibrium determines Π—and hence i—and Dk. In terms of Figure 2, this can be seen as follows. From given k, we can determine which $c'c'$ curve we are on. From given m, k, and Π^e we also know $\bar{\bar{m}}$, and hence we can determine where we are on the $c'c'$ curve. We can then read Dk off the vertical axis and determine Π by the mm curve passing through the point on the $c'c'$ curve. As k and Π change through time, the time path of the economy can be determined by the locus of the intersections of the appropriate mm and $c'c'$ curves. Steady state in the economy occurs, of course, at the origin in Figure 2, where $Dk=\bar{\bar{m}}=0$.

Now consider at time t_0 a perturbation of a steady state by a once-for-all increase in the rate of expansion of the nominal money stock from μ_0 to μ_1. The immediate effect of an increase in μ is to create an excess stock supply of money: by our static expectations hypothesis the demand for real balances falls instantaneously while the actual stock rises initially (i.e., before there is any effect on



(where $\pi_2 > \pi_1 > \pi^* > \pi_0$ and $k_1 > k^* > k_0$)

FIGURE 2

the price level). At t_0, $\bar{\bar{m}}$ will be positive and there will be an excess demand for goods as desired consumption rises and wealth holders try to dishoard by entering the goods market. This excess demand causes the rate of inflation to rise and the economy moves to a more southeasterly *mm* curve; initially the capital stock is given and we remain on the original $c'c'$ curve. The economy then has moved to point A in Figure 2 where Π has been bid up to Π_2 and Dk is less than zero.

At A, $\bar{\bar{m}}$ is positive, and due to our static expectations hypothesis, unambiguously falling.[14] On the other hand, $\bar{\bar{k}}$ is negative and falling ($Dk \leq 0$), causing $c'c'$ to shift northeasterly. Then, in Figure 2, the economy is moving northwesterly from A, and Π is falling. Our expectations hypothesis is sufficient to require that $\bar{\bar{m}}$ fall asymptotically to zero. The new steady state will be characterized by a higher capital intensity and hence the adjustment path will be constrained to satisfy

$$\int_{\tau A}^{T_0} Dk\mathrm{d}t = \theta$$

where θ is the change in the equilibrium capital intensity derived from the comparative dynamics of the model.[15] Two possible adjustment paths $ABCO$ and $ABDO$ are plotted; k, and hence m^d, attain their minimum values at B, then rise directly to their respective new long-run equilibrium values. Along $ABDO$, m and Π fall directly to their respective long-run equilibrium values, whereas along $ABCO$, both m and Π would overshoot and then approach their new equilibrium values from below.

Along the segment AB, the dishoarding is reducing $\bar{\bar{m}}$, but increasing the absolute value of $\bar{\bar{k}}$ (negative); both effects serving to increase the savings rate, and hence increase the share of current output available for capital accumulation. Northwest of B, capital accumulation becomes positive and $\bar{\bar{k}}$ rises; $\bar{\bar{m}}$ continues to fall and both effects are now working towards the restoration of portfolio equilibrium. Figure 3 plots representative time paths of the variables for the two alternative adjustment paths, where the starred variables represent the new equilibrium values corresponding to μ_1.[16] In terms of Figure 1, the economy moves initially to P, and then adjusts gradually, first in a southwesterly direction (capital intensity falling) until it crosses m^* after which it moves southeasterly to the new equilibrium position S.

It is useful to compare the behavior of the capital stock in this model to that obtained in Sidrauski's analysis in which long-run desired balances are always held, but expectations are adaptive. In Sidrauski's model, an increase in the rate of monetary expansion causes the capital stock to initially fall, and then to rise to its new steady-state level. Also, in that model, the increased real balances initially cause an increase in consumption, but in this case it is due to the fact that the increased balances are treated initially as an income transfer since the new higher equi-

[14] To see this, consider $D\bar{\bar{m}} = m(\mu - n - \Pi) - B(k, \Pi^e) \cdot Dk$, so for $D\bar{\bar{m}} < 0$, we require $\Pi > \mu - n - (B/m) \cdot Dk$, which from equation (11) occurs whenever $m > 0$. Hence we can draw the heavy black horizontal arrows in Figure 2.

[15] Our model retains the neoclassical equilibrium properties. See Levhari and Patinkin for a thorough treatment of the comparative dynamics propositions of the model. If we had considered the simple case of a once-for-all increase in the nominal money stock, we would again have an initial excess supply of real balances, and the resulting adjustment path would be similar to that considered in the text, constrained so that

$$\int_{\tau_A}^{T_0} Dk\mathrm{d}t = 0$$

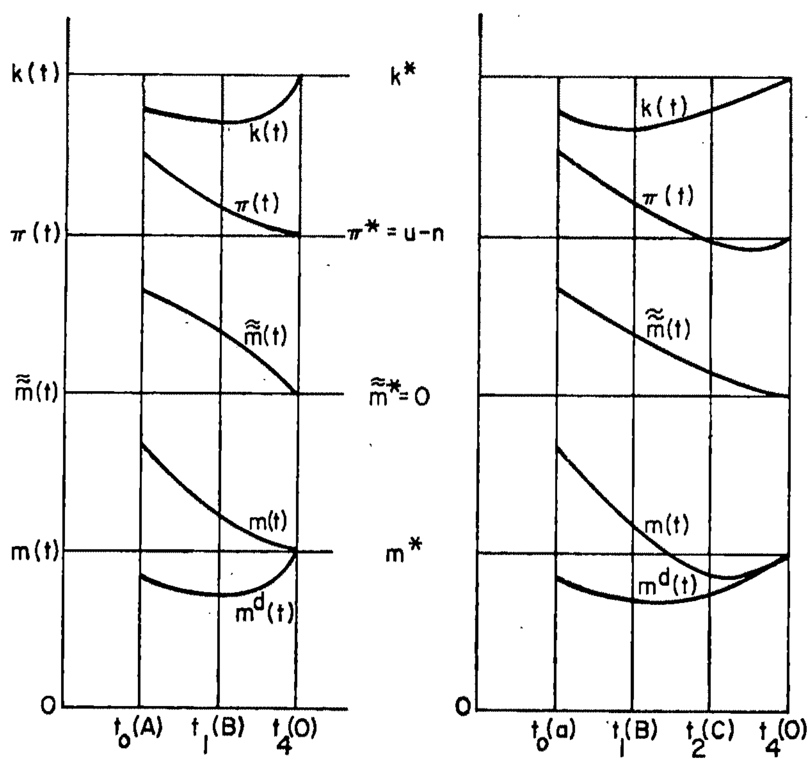
[16] The long-run effect on m^d is uncertain; the nominal rate of interest may rise or fall and hence, so may m^d (see Levhari-Patinkin, pp. 731–32). Figure 3 is drawn on the assumption that the steady-state value of $m^d(m^*)$ rises with μ.

librium rate of inflation is initially not anticipated. Hence, in a sense, there wealth holders also tend to consume out of their "excess stock supply of real balances." Only after a period of time sufficient for the increased rate of inflation to become more fully anticipated do they adjust their portfolios in favor of physical capital. If that model had our static expectations, the capital stock would rise immediately as the new steady-state inflation rate would be fully anticipated immediately.

In our present model, the actions of the monetary authority create a stock disequilibrium and the wealth holders then proceed to inject it into the economy over time by consuming out of their excess stock supply of real balances. The analogy to an economy with no adjustment costs but in which price expectations are sluggish, or in which the rate of expansion of the money supply is varied over time, is immediate.

Similar experiments such as money destruction, or national disasters which destroy part of the capital stock could also be treated, and the model could be easily extended to include such considerations as adaptive expectations and inside money.

While the present analysis still requires "blips" in the price level in response to monetary shocks, that seems to be a consequence of analyzing discrete policy jumps in a continuous time model. Our neoclassical variant does, however, appear capable of handling nonsteady-state behavior in a richer manner than the simple



3A ADJUSTMENT PATH ABDO

3B ADJUSTMENT PATH ABCO

FIGURE 3

neoclassical model which had virtually no explicit analysis of asset flows. The crucial feature of our model which yields this richer analysis is the explicit treatment given to the possibility of stock disequilibrium, and the ensuing analysis of the implications of such stock disequilibrium for asset flows and resulting interactions with the commodity market.

III. Concluding Comments

Finally consider the case suggested by Stein which might be construed as implying the following price equation:[17]

$$\begin{aligned}\Pi &= \mu - n + h \cdot \bar{\bar{f}} \\ (14) \qquad &= \mu - n + \frac{h}{1 + hm}\,[\bar{\bar{m}}(n + \alpha) \\ &\quad - B(k, \Pi^e) \cdot Dk]\end{aligned}$$

Again, in the steady state, $\bar{\bar{f}} = \bar{\bar{m}} = Dk = 0$, so equation (14) has the desired steady-state solution $\Pi = \mu - n$; and Dk can be related to $\bar{\bar{m}}$ and z by equation (15):

$$(15) \qquad Dk = -\frac{1}{B(k, \Pi^e)} \cdot \left[z\left(\frac{1}{h} + m\right) - \bar{\bar{m}}(n + \alpha) \right],$$

which is similar to equation (12). We also have the same conditions in the goods market, except we are not here constrained to flow equilibrium, and hence would conduct the analysis in terms of the cc curves, and not the $c'c'$ curves.

[17] Stein's analysis implies $\bar{\bar{f}} = q \cdot \bar{\bar{m}}$. Solving for q from equation (8):

$$\begin{aligned} q &= n - z\frac{m}{\bar{\bar{m}}} - \frac{B(k, \Pi^e)}{m} \cdot Dk + \alpha \\ &= q(\mu, k, Dk), \end{aligned}$$

so it is really not consistent with our formulation of the money flow demand given by equation (8) to treat q as a constant. Note that the limits of equation (14) and (15) as $h \to \infty$ are (11) and (12); this gives us the most restrictive case of all with only steady-state behavior possible, i.e., in Figure 1 the economy is restricted not only to positions along OR, but simply to be at E.

Thus, either equations (11) and (12), or equations (14) and (15) might be used to analyze nonsteady-state behavior. The first set represents essentially the neoclassical view wherein the price level is essentially driven by the money market, but in which costs of adjustment in asset markets which give rise to stock disequilibria phenomena are allowed for. The latter set is perhaps more in the spirit of the Keynes-Wicksell paradigm, inasmuch as nonsteady-state behavior in the price level requires nonzero excess flow demand for real balances, and therefore nonzero excess demand in the goods market. Then this latter model still requires some arbitrary method of allocating frustrations. A second problem arising in this latter model is that, just as in the original Keynes-Wicksell models, no forces operating to rationalize stock disequilibrium were introduced, and nothing is inherent in the model to prevent the achievement of flow equilibrium. This is particularly disturbing when our flow money demand equation (8) is introduced as we do here to obtain equation (14).

It therefore seems worthwhile treating in more detail the fact that in our neoclassical variant there need be no arbitrary allocation of frustrations. In terms of the demand relationships "effective" at a point in time, there need be no frustrations for there is perpetual flow equilibrium. Wealth holders are frustrated only in the sense that they do not instantaneously achieve their desired long-run wealth position. However, the wealth holders do make plans to adjust their wealth position in an optimal manner, and there is nothing inherent in the model to prevent the attainment of these plans.

It is hoped that the present paper contributes to the reconciliation sought by Stein in the concluding section of his paper. There he states, "The steady state must be the asymptotic solution of the short-run

dynamic model: It cannot be brought in as a *deus ex machina*" (p. 104). I believe that the neoclassical variant introduced in Section II of the present paper meets this criteria while maintaining the clearly desirable steady-state features of the neoclassical model, and introducing short-run characteristics which are more appealing and meaningful from the viewpoint of contemporary economic thought. Of course, as a description of short-run macro-economic behavior the present model is incomplete. A complete model would, at the very least, need to incorporate the effects of uncertainty and expectational disequilibrium into the analysis and allow for the existence of unemployment.

REFERENCES

S. Fischer, "Keynes-Wicksell and Neoclassical Models of Money and Growth," *Amer. Econ. Rev.*, Dec. 1972, *62*, 880–90.

J. P. Gould, "Adjustment Costs in the Theory of Investment of the Firm," *Rev. Econ. Stud.*, Jan. 1968, *35*, 47–55.

H. Grossman, "Money, Interest, and Prices in Market Disequilibrium," *J. Polit. Econ.*, Sept. 1971, *79*, 943–61.

H. G. Johnson, "The Neoclassical One-Sector Growth Model: A Geometrical Exposition and Extension to a Monetary Economy," *Economica*, Aug. 1966, *33*, 265–87.

D. Levhari and D. Patinkin, "The Role of Money in a Simple Growth Model," *Amer. Econ. Rev.*, Sept. 1968, *58*, 713–53.

K. Nagatani, "A Monetary Growth Model with Variable Employment," *J. Money, Credit, Banking*, May 1969, *1*, 228–51.

D. Purvis, "Introducing Useful Money into a Growth Model," *Can. J. Econ.*, Aug. 1971, *4*, 374–81.

H. Rose, "Unemployment in a Theory of Growth," *Int. Econ. Rev.*, Sept. 1966, *7*, 50–58.

M. Sidrauski, "Inflation and Economic Growth," *J. Polit. Econ.*, Dec. 1967, *75*, 796–810.

R. Solow, "A Contribution to the Theory of Economic Growth," *Quart. J. Econ.*, Feb. 1956, *70*, 796–810.

J. Stein, "Monetary Growth Theory in Perspective," *Amer. Econ. Rev.*, Mar. 1970, *60*, 85–106.

J. Tobin, "Money and Economic Growth," *Econometrica*, Oct. 1965, *33*, 671–84.

H. Uzawa, "Time Preference and the Penrose Effect in a Two-Class Model of Economic Growth," *J. Polit. Econ.*, Aug. 1969, *77*, 628–53.

An Analysis of Cooperation and Learning in a Duopoly Context

By RICHARD M. CYERT AND MORRIS H. DEGROOT*

A number of recent papers have introduced the notion of treating duopoly as a multiperiod maximization problem (see Cyert and DeGroot (1970a), and James Friedman (1968)). Previously, duopoly was viewed simply as a one-period maximization problem or as some minor variation of the latter. Recent work has shown that through the utilization of the concept of backward induction, models can be developed free from the earlier restrictions of limited period analysis.[1]

In all of these models, certain fixed assumptions about the behavior of the two firms are made, and optimal strategies are then developed for the firms in a multiperiod problem based on these fixed behavioral assumptions. In a problem with a large number of periods, an equilibrium position is ultimately established. There are no models currently in existence that enable the firms to change their behavioral assumptions and their strategies during the process of reaching an equilibrium position.

In this paper we develop a model in which the firms move in a rational fashion, from a path leading to a noncooperative equilibrium to paths leading to equilibria defined by various degrees of cooperation and ultimately to the cooperative (collusive) solution. In the process of constructing this model, we will introduce the notions of a coefficient of cooperation and of mutually optimal reaction functions. We will derive such mutually optimal reaction functions for various coefficients of cooperation. We will also utilize the concept of Bayesian learning contained in previous papers (see Cyert and DeGroot (1970b), (1971)).

I. Coefficient of Cooperation

Consider a duopoly problem in which the demand function in any given period is $p=f(q, r)$, where q is the output of firm 1 in that period and r is the output of firm 2. If $c_1(q)$ denotes the cost function of firm 1 in each period and $c_2(r)$ denotes the cost function of firm 2, then the profit π_i of firm i $(i=1, 2)$ will be

$$\pi_1(q, r) = qf(q, r) - c_1(q),$$
$$\pi_2(q, r) = rf(q, r) - c_2(r)$$

Thus in a process with n periods in which q_j and r_j denote the outputs in period j $(j=1, \ldots, n)$, the total profit of firm i over the n periods will be

$$\sum_{j=1}^{n} \pi_i(q_j, r_j)$$

The basic goal of each firm is to maximize its total profit. It is well known that this total profit will be higher for both firms

* Carnegie-Mellon University. This research was supported in part by the National Science Foundation under grants GP-23708 and GS-32514. We are indebted to James Friedman, Martin Geisel, Robert Lucas, Edward Prescott, Nancy Schwartz, and Oliver Williamson for their helpful comments on an earlier version of this paper. We also benefited greatly from the coments of the members of the NBER Seminar on Bayesian Inference in Econometrics held in Pittsburgh in May 1971.

[1] Duopoly is a particular type of nonzero sum, two-person game which is related to, but distinct from, the type known as the Prisoner's Dilemma. Robert Luce and Howard Raiffa (pp. 97–102) have discussed a multiperiod analysis of the Prisoner's Dilemma, and F. Trenery Dolbear et al. (1968, 1969) have pointed out the connections between duopoly and the Prisoner's Dilemma.

when they cooperate with each other than it will be when each firm attempts to maximize its own profit without regard to its rival. We propose, therefore, to introduce a utility function for each firm that will encompass the degree of cooperation the firm is prepared to demonstrate toward its rival.

We shall assume that there exists a constant $\gamma (0 \leq \gamma \leq 1)$ such that firm 1 desires to maximize the sum

$$(1) \quad V_1 = \sum_{j=1}^{n} [\pi_1(q_j, r_j) + \gamma \pi_2(q_j, r_j)]$$

In other words, it is assumed that V_1 represents the utility function for firm 1. If $\gamma = 0$, then the sum in (1) is simply the total profit of firm 1 over the n periods. However, if $\gamma \neq 0$, then the sum in (1) is a linear combination of the total profit of firm 1 and that of firm 2. The number γ will be called the *coefficient of cooperation* of firm 1. If $0 < \gamma < 1$, firm 1 is interested in cooperating with firm 2 and desires to maximize its own total profits plus a fraction of the profits of firm 2. If $\gamma = 1$, firm 1 is interested in cooperating with firm 2 to the extent of desiring simply to maximize the total profits of the two firms.[2] Although values of γ outside the interval $0 \leq \gamma \leq 1$ are conceptually possible, such values do not seem to be of practical interest and will not be considered further here.

Similarly, we shall assume that there also exists a constant $\delta (0 \leq \delta \leq 1)$ such that firm 2 desires to maximize a utility function V_2 of the following form:

$$(2) \quad V_2 = \sum_{j=1}^{n} [\pi_2(q_j, r_j) + \delta \pi_1(q_j, r_j)]$$

Thus, δ is the coefficient of cooperation of firm 2.

[2] Friedman (1962) refers to a similar function and credits G. Feeney with suggesting it. However, Friedman does not make the same use of the concept as has been done in this paper, nor has Feeney to the best of our knowledge.

It should be emphasized that in the theory to be presented here, there is a rational reason why each firm attempts to maximize a linear combination such as V_1 or V_2 rather than simply trying to maximize its own profits. The firms recognize that for certain pairs of positive values of γ and δ, their profits will actually be larger when they try to maximize V_1 and V_2 than when each firm tries to maximize its own profits directly. We have called V_1 and V_2 the utility functions of the firms because they choose their outputs in order to maximize these functions. This terminology is slightly inappropriate, however, since the firms ultimately evaluate the usefulness of given coefficients γ and δ, not in terms of how large the values of V_1 and V_2 are, but in terms of how large their individual profits are.

II. Mutually Optimal Reaction Functions

Ignoring the values of the coefficients of cooperation for the moment, we define two reaction functions to be mutually optimal if the intersection point of the two reaction functions is also the point of maximum profit for each firm relative to the reaction function of the rival. This concept can be made clearer by looking at the traditional Cournot model where the reaction functions are not mutually optimal. If firm 1 knows that firm 2 is using the Cournot assumptions, then firm 1 can use this information to develop a new reaction function that will intersect the reaction function of firm 2 at a point that is more profitable for firm 1 (see R. G. D. Allen, pp. 345–47). On the other hand, two reaction functions, one vertical and one horizontal that intersect at the Cournot equilibrium point are mutually optimal. With this preface we undertake a more complete definition and analysis.

Consider now a duopoly problem in which the number of periods n is large. We shall assume that the two firms choose

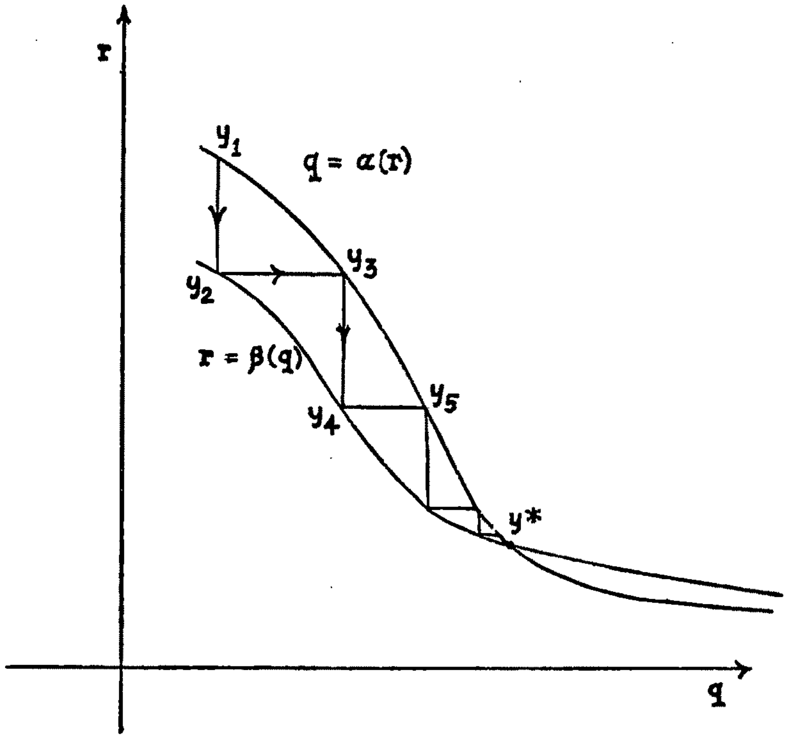


FIGURE 1

their outputs in alternating periods, that firm 1 chooses its output q according to a reaction function $q=\alpha(r)$, and that firm 2 chooses its output r according to a reaction function $r=\beta(q)$. In other words, if firm 2 has just chosen its output in a given period to be r_0, then firm 1 will choose its output in the *next* period to be $q_1=\alpha(r_0)$ while firm 2 holds its output at r_0. In the following period, firm 2 will then choose its output to be $r_2=\beta(q_1)$ while firm 1 holds its output at q_1. The choices continue to be made in this alternating fashion through successive periods.

If the curves $q=\alpha(r)$ and $r=\beta(q)$ intersect at the point $y^*=(q^*, r^*)$, then as indicated in Figure 1, the sequence of successive points

$$y_1 = (q_1, r_0), \quad y_2 = (q_1, r_2),$$
$$y_3 = (q_3, r_2), \quad y_4 = (q_3, r_4), \ldots$$

will typically converge to the point y^*. This convergence to the point of intersection y^* will take place for any reaction functions α and β satisfying the following three properties:[3]

1. Both the curves $q=\alpha(r)$ and $r=\beta(q)$ are decreasing (or nonincreasing).

2. The curve $q=\alpha(r)$ lies above the curve $r=\beta(q)$ for all points (q, r) such that $q<q^*$.

3. The curve $q=\alpha(r)$ lies below the curve $r=\beta(q)$ for all points (q, r) such that $q>q^*$.

Thus, under these conditions, if the two firms repeatedly use the reaction functions $q=\alpha(r)$ and $r=\beta(q)$, their outputs will

[3] These conditions should be familiar to the reader since they are analogous to the constraints on the slopes of the marginal revenue and marginal cost curves in standard competitive theory (see Paul Samuelson, pp. 57–89).

converge to q^* and r^*, regardless of the initial outputs at which the process starts.

Suppose now that the coefficients of cooperation of firms 1 and 2 are γ and δ, respectively, and that each firm knows the rival's cost and reaction functions. In any given period, firm 1 knows that corresponding to any value that it chooses for its output q, firm 2 will react in the next period by choosing the value $r=\beta(q)$. The profits of firms 1 and 2 will then be:

$$\pi_1[q, \beta(q)] = qf[q, \beta(q)] - c_1(q)$$

and

$$\pi_2[q, \beta(q)] = \beta(q)f[q, \beta(q)] - c_2[\beta(q)]$$

Since the coefficient of cooperation of firm 1 is γ, firm 1 is interested in maximizing the value of the utility

$$P_1 = \pi_1[q, \beta(q)] + \gamma\pi_2[q, \beta(q)] \tag{3}$$

Suppose that this utility is maximized at the value $q=q_M$. Because of the convergence which we have described, we know that in the long run, the output of firm 1 will be close to the value $q=q^*$ in every period. It follows, therefore, that from the point of view of firm 1, the most desirable value of q^* is $q^*=q_M$.

For this reason, we shall say that the reaction function $q=\alpha(r)$ of firm 1 *is optimal against* the reaction function $r=\beta(q)$ of firm 2 if $q^*=q_M$. In other words, the reaction function $q=\alpha(r)$ of firm 1 is optimal against the reaction function $r=\beta(q)$ of firm 2 if the point of intersection (q^*, r^*) of the two reaction functions, toward which the outputs of the two firms converge, is also the point on the curve $r=\beta(q)$ at which the utility P_1 of firm 1 is a maximum.

Similarly, for any choice of output r by firm 2 the reaction of firm 1 will be $q=\alpha(r)$ and the profits of the two firms will be $\pi_1[\alpha(r), r]$ and $\pi_2[\alpha(r), r]$. The utility function of firm 2 is

$$P_2 = \pi_2[\alpha(r), r] + \delta\pi_1[\alpha(r), r] \tag{4}$$

and we shall let r_M denote the value of r for which this utility is maximized. Then we say that the reaction function $r=\beta(q)$ of firm 2 is optimal against the reaction function $q=\alpha(r)$ of firm 1 if $r^*=r_M$.

Finally, we say that the two reaction functions are mutually optimal if both $q_M=q^*$ and $r_M=r^*$. In other words, the two reaction functions are mutually optimal if the preferred position of firm 1 along the curve $r=\beta(q)$ is at the point (q^*, r^*) and the preferred position of firm 2 along the curve $q=\alpha(r)$ is also at the point (q^*, r^*). It should be emphasized that the coefficients of cooperation γ and δ are assumed to have given values in this definition and that two reaction functions are considered to be mutually optimal *relative to the given values* of γ *and* δ.

When the two reaction functions are mutually optimal, neither firm will desire to change the point of intersection toward which the outputs are converging by changing its own reaction function as long as the rival firm does not change.

III. Linear Reaction Functions

We shall now apply these concepts to the special but important case in which the demand function is linear and the cost functions of both firms are 0. As described in our previous paper, we can assume without loss of generality that the demand function is of the form $F(q, r)=1-q-r$,[4] so that the profit functions for the firms are

$$\begin{aligned} \pi_1(q, r) &= q - q^2 - qr \\ \pi_2(q, r) &= r - qr - r^2 \end{aligned} \tag{5}$$

Suppose also that both the reaction function $q=\alpha(r)$ of firm 1 and the reaction function $r=\beta(q)$ of firm 2 are linear functions specified as follows:

[4] For simplicity we have not constrained $(q+r)$ to be less than 1 in this form. Thus it is possible to get negative outputs as in Table 2. A negative output stemming from the manipulation of the model has no economic significance.

$$\alpha(r) = T_1 + U_1 r, \tag{6}$$
$$\beta(q) = T_2 + U_2 q,$$

where $|U_i| < 1$ for $i = 1, 2$. Finally, assume that the coefficient of cooperation of firm 1 is γ and that of firm 2 is δ.

For $0 \leq \gamma \leq 1$, the value q_M that maximizes the utility P_1 of firm 1, as given by equation (3), can now be found by elementary differentiation to be

$$q_M = \frac{1 - T_2 + \gamma(U_2 - T_2 - 2T_2U_2)}{2(1 + \gamma U_2)(1 + U_2)} \tag{7}$$

Similarly, for $0 \leq \delta \leq 1$, the value r_M that maximizes the utility P_2 of firm 2, as given by equation (4) will be

$$r_M = \frac{1 - T_1 + \delta(U_1 - T_1 - 2T_1U_1)}{2(1 + \delta U_1)(1 + U_1)} \tag{8}$$

Furthermore, the coordinates of the point of intersection of the reaction functions specified by (6) are

$$q^* = \frac{T_1 + T_2U_1}{1 - U_1U_2} \tag{9}$$

and

$$r^* = \frac{T_2 + T_1U_2}{1 - U_1U_2} \tag{10}$$

The reaction functions specified by (6) will be mutually optimal if, and only if, $q_M = q^*$ and $r_M = r^*$.

For any given values of γ and δ, there will be an infinite number of pairs of reaction functions that are mutually optimal. In fact, a pair of reaction functions specified by (6) will be mutually optimal if, and only if, the four numbers T_1, U_1, T_2, and U_2 satisfy the two constraints $q_M = q^*$ and $r_M = r^*$.

For example, consider the traditional case in which $\gamma = \delta = 0$ and each firm attempts simply to maximize its own profits. In this case it follows from equations (7) to (10) that the reaction functions given in (6) will be mutually optimal if, and only if,

$$\frac{1 - T_2}{2(1 + U_2)} = \frac{T_1 + T_2U_1}{1 - U_1U_2} \tag{11}$$

and

$$\frac{1 - T_1}{2(1 + U_1)} = \frac{T_2 + T_1U_2}{1 - U_1U_2} \tag{12}$$

After simplification, equations (11) and (12) can be rewritten as follows:

$$T_1 = T_2 = \frac{1 - U_1U_2}{(2 + U_1)(2 + U_2) - 1} \tag{13}$$

It can be seen from equation (13) that for any choice of the slopes U_1 and U_2 such that $|U_i| < 1$ $(i = 1, 2)$, there is a pair of mutually optimal reaction functions having those slopes. Furthermore, it is seen from equation (13) that the intercepts T_1 and T_2 must be equal in every pair of mutually optimal reaction functions.

The profits of the two firms will, of course, be different for different pairs of mutually optimal reaction functions. In particular, for a pair such that $U_1 = U_2 = U$, the point of intersection will be $q^* = r^* = 1/3 + U$ and the profit for each firm at this point will be

$$\pi_1 = \pi_2 = \frac{1 + U}{(3 + U)^2} \tag{14}$$

It follows from equation (14) that as U increases from -1 to $+1$, the profit for each firm increases from 0 to $\frac{1}{8}$, which is the profit that they can obtain by dividing the monopoly profits between them equally.

When $U = 0$, the mutually optimal reaction functions specify that both firms will repeatedly choose their outputs at the Cournot equilibrium point $q = r = \frac{1}{3}$. It should be noted that the usual Cournot reaction functions $q = \frac{1}{2}(1 - r)$ and $r = \frac{1}{2}(1 - q)$ are not mutually optimal even though their point of intersection is also $q = r = \frac{1}{3}$.

IV. Optimal Strategies in Multiperiod Decision Problems with Alternating Choice

We shall now consider again a process with n periods and shall assume that the two firms choose their outputs in alternating periods as described earlier in this paper. In our previous paper we developed pure duopoly strategies for the two firms by using the method of backward induction based on the assumption that each firm desired to maximize its total profit over the n periods. Here we shall generalize this development to a problem in which the coefficients of cooperation γ and δ of the two firms are fixed, firm 1 desires to maximize the utility V_1 given by equation (1), and firm 2 desires to maximize the utility V_2 given by equation (2).

For each different pair of values of γ and δ, the method of backward induction specifies a different sequence of outputs for the two firms which, in turn, yield a different sequence of profits for the firms over the n periods. Thus, the choice of γ and δ by the firms fixes their profit stream over the n periods. Suppose, for example, that $\gamma=\delta$, so the firms are using the same coefficient of cooperation, and suppose that c_1 and c_2 are numbers such that $0 \leq c_1 < c_2 < 1$. Then, in general, the profits of both firms will be higher when $\gamma=\delta=c_2$ than they will be when $\gamma=\delta=c_1$. In other words, it will be mutually profitable for the firms to use a common coefficient of cooperation that is as large as possible. The development given in our previous paper (see Cyert and DeGroot (1970a)) corresponds to the problem in which $\gamma=\delta=0$, and the development to be sketched here for arbitrary values of γ and δ is completely analogous to the development given there.

We shall assume for convenience that the total number of periods n is even and that firm 2 must choose its output r_n in the final period n. After the output q_{n-1} of firm 1 in period $n-1$ is known, firm 2 will choose r_n to maximize its utility $\pi_2(q_{n-1}, r_n) + \delta\pi_1(q_{n-1}, r_n)$. Thus, when firm 1 chooses its output q_{n-1} in period $n-1$, it knows the output r_{n-2} of firm 2 in period $n-2$ as well as the reaction function that firm 2 will use in period n. Hence, subject to these conditions, it can choose q_{n-1} to maximize its utility over the final two periods

$$[\pi_1(q_{n-1}, r_{n-2}) + \gamma\pi_2(q_{n-1}, r_{n-2})] + [\pi_1(q_{n-1}, r_n) + \gamma\pi_2(q_{n-1}, r_n)]$$

By moving backward through the process in this way, it is seen that at any given period, the firm which must choose its output in that period will know the reaction functions that will be used by the rival firm and by itself in all later periods. Hence, for any given output of the rival in the preceding period, it can choose its own output to maximize the sum of its utilities over all the remaining periods. The sequences of reaction functions for the two firms constructed in this way are called the *pure duopoly strategies relative to the given coefficients of cooperation* γ *and* δ.

If this process of backward induction is continued indefinitely and the reaction functions for the firms are derived for earlier and earlier periods, then these reaction functions will typically converge to certain limiting reaction functions. These limiting reaction functions are presented in the Appendix for the problem in which the demand function is linear and the profit functions of the two firms are given by (5).[5] In this problem, the reaction function used in each period and the limiting reaction functions of both firms are linear functions.

It is shown in the Appendix that when

[5] The relations (A1) to (A13) given in the Appendix which govern the backward induction process and the limiting reaction functions, include a discount factor as discussed later in this paper. These relations can be applied to the problem being considered here, in which the profits of the firms over the n periods are not discounted, by letting $\rho=1$.

the profit functions of the two firms are given by (5) and the coefficients of cooperation γ and δ are fixed, the limiting reaction functions are mutually optimal. More generally, as described in our previous paper, a wide class of quadratic profit functions can be reduced to the form (5) by an appropriate choice of the units of measurement (see Cyert and DeGroot (1970a)). In particular, it follows, therefore, that in any problem in which the demand function is linear and the costs of the two firms are assumed to be zero, the limiting reaction functions will be mutually optimal.

Because of the nature of the process by which the pure duopoly strategies are derived, we believe that for a wide class of profit functions other than quadratic ones, it will remain true that the limiting reaction functions are mutually optimal. However, at the present time this property has been established only for the quadratic functions described here.

V. Changing Coefficients of Cooperation

Up to this point we have established the following conclusions:

1. For each fixed pair of coefficients of cooperation γ and δ, there exists an equilibrium point obtained from mutually optimal limiting reaction functions.

2. There exists a unique pair of paths leading to each such equilibrium point established by a process of backward induction in which each firm wants to be in the optimal position vis-à-vis its competitor at the final point.

3. When $\gamma=\delta=0$, the equilibrium point is the noncooperative solution obtained in our previous paper. When $\gamma=\delta$ and their common value increases toward the value 1, the equilibrium point yields higher profits for both firms.

These conclusions mean that each firm can be viewed as facing a decision set consisting of an infinite number of pairs of multiperiod decision paths. Each of these pairs of paths leads to an equilibrium point established from mutually optimal reaction functions. Each pair is identified by the values of the coefficients of cooperation held by each firm. On each path the alternating decisions assumption holds. In each decision period the firm making the decision in that period makes the optimum decision, given the rival's decision in the previous period, for maximizing its utility over the infinite horizon. This entire process is, of course, based on the assumption that neither demand nor cost conditions are changing.

It is not however, necessary to assume that the coefficients of cooperation are fixed throughout the entire process. Each firm can change its coefficient of cooperation at any time. Hence, no firm need stay on a fixed path.

In order to illustrate the model's potential, let us start with a situation in which both firms are on the paths consistent with the values $\gamma=\delta=0$. Assume that through the process of learning, one of the firms develops a probability distribution on the rival's likelihood of cooperating with a larger value for the coefficient of cooperation. Specifically, let firm 1 have a uniform prior distribution for firm 2's coefficient of cooperation δ on the interval $0\leq\delta\leq d$, where 0 is the current value of δ and d is the highest level of cooperation to which firm 1 believes firm 2 will implicitly consent. Let us further assume that if d is less than the value of γ that firm 1 uses, then firm 2 will keep $\delta=0$, and that if $\gamma\leq d$ then firm 2 will match its coefficient of cooperation and choose $\delta=\gamma$.

If we assume a two-period analysis, then it is possible to determine an optimal value for γ, the value that firm 1 should set. Let $W_1(\gamma\,|\,0, d)$ stand for the expected utility of firm 1 over the two periods for a given value of γ $(0\leq\gamma\leq d)$. The expression

for W_1 must take into account the fact that decisions are made in alternating periods. We assume a process in which each firm is proceeding along the optimum paths for $\gamma=\delta=0$. In the period appropriate for firm 1 to make its output decision, it does so on the basis of the previous output r_0 of firm 2, but with a new choice of γ. By choosing an output q^* that is consistent with r_0 and a new value of $\gamma>0$, firm 1 indicates to firm 2 that it is prepared to operate at a higher level of cooperation than previously. Firm 2 then has the opportunity to respond positively, that is, to choose an r_1^* based on q^* and $\delta=\gamma>0$, or to continue to hold $\delta=0$ and choose an r_2^* based on $\delta=0$ and q^*. On this basis, firm 1 must choose γ to maximize the following expected utility:

$$W_1(\gamma \mid 0, d) = [\pi_1(q^*, r_0) + \gamma\pi_2(q^*, r_0)] + \frac{d-\gamma}{d}[\pi_1(q^*, r_1^*) + \gamma\pi_2(q^*, r_1^*)] + \frac{\gamma}{d}[\pi_1(q^*, r_2^*) + \gamma\pi_2(q^*, r_2^*)]$$

Since q^*, r_1^*, and r_2^* are also functions of γ, the maximization process becomes messy although straightforward. The question is whether there is a value of $\gamma>0$ such that $W_1(\gamma|0, d)>W_1(0|0, d)$. Some economic insight can be gained by regrouping the terms in $W_1(\gamma|0, d)$ and examining the resulting expression. We have

$$W_1(\gamma \mid 0, d) = [\pi_1(q^*, r_0) + \pi_1(q^*, r_1^*)] + \gamma[\pi_2(q^*, r_0) + \pi_2(q^*, r_1^*)] + \frac{\gamma}{d}\left\{[\pi_1(q^*, r_2^*) - \pi_1(q^*, r_1^*)] + \frac{\gamma^2}{d}[\pi_2(q^*, r_2^*) - \pi_2(v^*, r_1^*)]\right\}$$

It is reasonable to assume that the first two expressions are positive. The third one should be negative since we would expect firm 1 to make more profit when firm 2 cooperated by making $\delta=\gamma$. The fourth term should be positive since we would expect firm 2 to make more profit by being uncooperative (while firm 1 is being cooperative) than by being cooperative. Thus, the impact of firm 2's noncooperative action on firm 1 becomes decisive in determining whether the expected utility is increasing and in determining the size of the optimum γ if W_1 is positive. The larger the absolute value of the third term, the smaller the optimum value of γ will be. This result, of course, is not surprising since it is the response of firm 2 to firm 1's action that is the uncertain element in the situation.

Though this analysis is for the two-period case, it illustrates the process by which the firms can move from one track leading to a mutually optimal equilibrium to another. It is possible in this way for the firm to move from the uncooperative track, $\gamma=\delta=0$, to the higher profits corresponding to $\gamma=\delta=1$. Once learning takes place and the optimal value of γ becomes $\gamma^*>0$, firm 1 can rationally attempt to move to the new track, namely the one for which both $\gamma=\delta=\gamma^*$. Firm 2 then has two alternatives. It can refuse to follow and make larger profits than it did previously. However, it can expect firm 1 to come back eventually to the track represented by $\gamma=\delta=0$. Under such a situation firm 2 is back to its original profit position. If firm 2, on the other hand, moves to the new track by making $\delta=\gamma^*$, its profits will be greater than they were for $\delta=\gamma=0$. With the usual assumptions about demand and cost curves it is reasonable to assume that profits and hence utility become greater as $\gamma=\delta\rightarrow 1$.

It is possible to extend the analysis for the optimum new coefficient of cooperation to more periods than the two assumed. It is possible to view the expected utility over n periods and to solve for the optimum decision horizon as well as the

optimum coefficient of cooperation. It is also possible to postulate a different decision rule for the firm initiating the move to the higher level of cooperation. We assumed that firm 1 determined its output q^* based on the γ and the current output of firm 2, r_0. Many other decision rules might have been used by firm 1. The main point is that firm 1 wants to behave in such a way that firm 2 clearly understands that firm 1 is signalling a change in the level of cooperation.

The process just described continues to occur over time. If firm 1's efforts succeed and firm 2 increases its δ to make it equal to γ, then both firms are on a track leading toward a new mutually optimal equilibrium point. Both firms are making more profits than they were when each was on the uncooperative track. Both firms continue to learn and form priors about the new values toward which the other firm will raise its coefficient of cooperation. At some point when the decision rule being used shows that the expected gain in utility is worth the risk of increasing the coefficient of cooperation, one of the firms will make the move and eventually the firms will move to another pair of tracks where the profit is higher. In this model either firm may initiate the movement.

Even if the first move should fail to stimulate the rival to match the new coefficient of cooperation, eventually the process will succeed. The temporary gains from not matching will be lost as soon as the firm initiating the upward movement reduces its coefficient of cooperation back to the old. Punishment is as effective a stimulus for learning for real firms in the real world as it is for subjects in an experiment in a laboratory.

We argue, therefore, that this model demonstrates the kind of behavior that many economists have speculated does occur in duopoly and oligopoly. We have shown how firms progress from a noncooperative solution to situations of increasing profit and ultimately to the tracks characterized by $\gamma=\delta=1$.

The dynamicization of the duopoly model has been accomplished by the introduction of the concept of firm learning, by the formalization of the learning through Bayesian analysis, and by utilizing the concept of mutually optimal equilibria all within the format of multiperiod analysis. We believe this pattern of concepts can also be used to extend and improve other duopoly and oligopoly models (see Cyert and DeGroot (1971)).

VI. Discounting

Up to this point no mention has been made of the discounting process. We have made our analysis with the discount rate equal to zero. Such an assumption is, of course, unrealistic.

We shall continue to consider an n period problem in which the coefficients of cooperation of the two firms are γ and δ. We shall assume, however, that instead of desiring to maximize the sum (1), firm 1 desires to maximize the following discounted sum

$$(15)\quad V_1 = \sum_{i=1}^{n} \rho^{i-1}[\pi_1(q_i, r_i) + \gamma\pi_2(q_i, r_i)],$$

where ρ is the discount factor and is equal to $1/(1+r)$, where r is the discount rate. Thus $\rho=1$ is equivalent to no discounting ($r=0$) and $\rho=0$ implies $r=\infty$. Similarly, it is assumed that firm 2 desires to maximize the discounted sum

$$(16)\quad V_2 = \sum_{i=1}^{n} \rho^{i-1}[\pi_2(q_i, r_i) + \delta\pi_1(q_i, r_i)]$$

When the firms choose their outputs in alternating periods, the optimal sequence of reaction functions can again be found by backward induction. As before, if the process of backward induction is continued indefinitely and the reaction functions are derived for earlier and earlier

periods, these reaction functions will typically converge to certain limiting functions.

These limits are presented in Table 1 for various values of γ and δ (with $\gamma=\delta$) and various values of ρ, for the simple model for which the profit functions π_1 and π_2 of the two firms are given by (5). The relations governing the backward induction process in this model are presented in the Appendix.

It follows from the relations given in the Appendix, and it can be seen in Table 1, that for a discount factor of 0 ($r=\infty$), the solution for a coefficient of cooperation of 0 is the Cournot solution. However, as the discount factor increases (r decreases) so that the future is weighted more heavily, a zero coefficient of cooperation does not lead to the Cournot solution. In fact, for higher discount factors the Cournot output and profit (but not the Cournot reaction functions) can be found at higher coefficients of cooperation. This fact reemphasizes the point made in our earlier paper (see Cyert and DeGroot (1970a)) that the Cournot solution requires a certain amount of trust. Thus for the true noncooperative solution (coefficient of cooperation $=0$) the profit continues to fall as the discount factor increases. For any given discount factor $\rho<1$ it can be seen that the firms move toward the monopoly solution as the coefficient of cooperation increases and attain it when the coefficient is equal to one.

It should be pointed out, however, that when a discount factor is present, the

TABLE 1—LIMITING VALUES FOR VARIOUS DISCOUNT FACTORS WHEN $\gamma=\delta$

ρ	$\gamma=\delta$	$T_1=T_2$	$U_1=U_2$	$q^*=r^*$	$\pi_1=\pi_2$
.95	0	.4805	−.3031	.3687	.0968
	.2	.4953	−.3931	.3556	.1027
	.5	.5346	−.5719	.3401	.1088
	.9	.6133	−.9185	.3197	.1153
	1.0	.5000	−1.0000	.2500	.1250
.8	0	.4848	−.3275	.3652	.0984
	.2	.4994	−.4238	.3508	.1047
	.5	.5350	−.6122	.3318	.1116
	.9	.5601	−.9381	.2890	.1220
	1.0	.5000	−1.0000	.2500	.1250
.5	0	.4931	−.3844	.3562	.1025
	.2	.5047	−.4901	.3387	.1093
	.5	.5256	−.6787	.3131	.1170
	.9	.5198	−.9459	.2671	.1244
	1.0	.5000	−1.0000	.2500	.1250
.2	0	.4988	−.4513	.3437	.1075
	.2	.5038	−.5571	.3236	.1142
	.5	.5100	−.7253	.2956	.1208
	.9	.5053	−.9486	.2595	.1248
	1.0	.5000	−1.0000	.2500	.1250
0	0	.5000	−.5000	.3333	.1111
	.2	.5000	−.6000	.3125	.1172
	.5	.5000	−.7500	.2857	.1224
	.9	.5000	−.9500	.2564	.1249
	1.0	.5000	−1.0000	.2500	.1250

TABLE 2—LIMITING VALUES FOR VARIOUS COEFFICIENTS OF COOPERATION WHEN THERE IS NO DISCOUNTING

γ	δ	T_1	U_1	T_2	U_2	q^*	r^*	π_1	π_2
0	0	.4791	−.2956	.4791	−.2956	.3698	.3698	.0963	.0963
0	.1	.5001	−.3023	.4641	−.3295	.3996	.3325	.1071	.0891
0	.2	.5244	−.3096	.4472	−.3646	.4350	.2886	.1202	.0798
0	.3	.5527	−.3178	.4277	−.4009	.4777	.2362	.1367	.0676
0	.4	.5866	−.3269	.4049	−.4389	.5303	.1721	.1578	.0512
0	.5	.6279	−.3372	.3776	−.4786	.5969	.0919	.1857	.0286
0	.6	.6799	−.3491	.3440	−.5205	.6841	−.0121	.2244	−.0040
0	.7	.7482	−.3632	.3010	−.5652	.8039	−.1534	.2810	−.0536
0	.8	.8435	−.3802	.2425	−.6135	.9799	−.3587	.3712	−.1359
0	.9	.9903	−.4020	.1552	−.6668	1.2677	−.6901	.5355	−.2915
.2	.2	.4938	−.3834	.4938	−.3834	.3570	.3570	.1021	.1021
.2	.3	.5229	−.3942	.4750	−.4226	.4027	.3048	.1178	.0891
.2	.4	.5574	−.4062	.4529	−.4636	.4601	.2395	.1382	.0719
.2	.5	.5996	−.4196	.4262	−.5069	.5344	.1553	.1658	.0482
.2	.6	.6529	−.4350	.3929	−.5527	.6345	.0422	.2051	.0136
.2	.7	.7237	−.4531	.3496	−.6020	.7773	−.1183	.2651	−.0403
.2	.8	.8251	−.4752	.2891	−.6556	.9989	−.3658	.3665	−.1342
.2	.9	.9901	−.5038	.1940	−.7159	1.3957	−.8051	.5714	−.3296
.5	.5	.5337	−.5580	.5337	−.5580	.3425	.3425	.1079	.1079
.5	.6	.5872	−.5787	.5003	−.6094	.4598	.2201	.1472	.0705
.5	.7	.6585	−.6022	.4561	−.6645	.6399	.0309	.2107	.0102
.5	.8	.7630	−.6303	.3926	−.7246	.9490	−.2951	.3284	−.1021
.5	.9	.9465	−.6664	.2856	−.7924	1.6022	−.9840	.6117	−.3757
.8	.8	.6038	−.8042	.6038	−.8042	.3346	.3346	.1107	.1107
.8	.9	.7760	−.8382	.4790	−.8710	1.3872	−.7293	.4745	−.2494
.9	.9	.6341	−.9005	.6341	−.9005	.3337	.3337	.1110	.1110

limiting reaction functions obtained from the backward induction process will not be mutually optimal.

VII. Summary and Conclusions

This paper presents an extensive generalization of the work that Cournot began in 1839. We have generalized the number of periods involved in the analysis. We have introduced the concept of the coefficient of cooperation and have shown that by varying the discount factor ρ and the values of the coefficients of cooperation γ and δ an infinite number of equilibrium positions are possible. Only one combination, $\rho=0$ and $\gamma=\delta=0$, gives the classical Cournot solution; although the Cournot equilibrium values are shown to exist for $\gamma=\delta>0$ for $\rho<1$. In fact, for any value of ρ there is a value of γ and δ that will give the Cournot values.

Cournot had to restrict his firms to a single assumption about their rival's behavior. By the introduction of the concept of learning and by making the management of each of the firms Bayesians, we are not required to restrict our firms to a single position with respect to their rival during the course of the analysis. We also free the analysis of the traditional criticism against Cournot, i.e., the management of each firm is stupid and does not learn even though the rival firm repeatedly behaves contrary to assumption (see George Stigler). In our model the firm can learn more about the nature of his rival and this learning affects the value of the coefficient of cooperation set by the firm. In the model developed in this paper, either firm may function as the leader in trying to move the two firms to a higher profit track represented by an increased

value of the coefficient of cooperation. The decision is made (or not made) to move to a higher track by one of the firms on the basis of a probability distribution developed through learning applied to expected profits. We demonstrate that the process of learning with changing priors on the rival's behavior enables the firms to move to the monopoly solution, given that $\rho<1$. We also show that for $\rho=1$ and for $0\leq\gamma\leq1$ and $0\leq\delta\leq1$, each track for the firms results in a mutually optimal equilibrium position. This result does not hold when $\rho<1$, in other words, when the future is discounted.

Appendix

We shall derive here the optimal strategies for each firm when the profit functions in each period are given by (5), the coefficients of cooperation are γ and δ, and firms desire to maximize the discounted sums V_1 and V_2 given by (15) and (16). When $\rho=1$, equations (15) and (16) reduce to equations (1) and (2), so the discussion which follows includes as a special case the problem in which there is no discounting ($\rho=1$).

As far as possible, we shall follow the notation of our previous paper (see Cyert and DeGroot (1970a)). It is assumed that the total number of periods n is even, that firm 1 must choose its output q_j in each odd period j, and that firm 2 must choose its output r_k in each even period k.

By backward induction the optimal choice of firm 1 in period j can be shown to be a linear function of the output r_{j-1} of firm 2 in period $j-1$. We shall denote this optimal choice by

$$q_j^* = t_j + u_j r_{j-1} \tag{A1}$$

When firm 1 chooses the output q_j^* and all subsequent choices by the two firms on periods $j+1$ to n are also made optimally, it can be shown by backward induction that the sum of the utilities for both firm 1 and firm 2 on periods j to n will be quadratic functions of r_{j-1}. If $F_j(r_{j-1})$ denotes the sum of these utilities for firm 1 and $G_j(r_{j-1})$ denotes the sum for firm 2, then these sums are of the form:

$$F_j(r_{j-1}) = A_{j-1} + B_{j-1}r_{j-1} + C_{j-1}r_{j-1}^2 \tag{A2}$$

$$G_j(r_{j-1}) = D_{j-1} + H_{j-1}r_{j-1} + L_{j-1}r_{j-1}^2 \tag{A3}$$

Similarly, firm 2 must choose the output r_k in each even period k. The optimal value of r_k will be a linear function of q_{k-1} which we shall denote

$$r_k^* = t_k + u_k q_{k-1} \tag{A4}$$

With this choice of r_k and optimal choices of outputs by the two firms on all subsequent periods from $k+1$ to n, the sum of the utilities for firm 2 on periods k to n, as a function of q_{k-1}, will be denoted

$$G_k(q_{k-1}) = A_{k-1} + B_{k-1}q_{k-1} + C_{k-1}q_{k-1}^2 \tag{A5}$$

The sum of the utilities for firm 1 will be denoted

$$F_k(q_{k-1}) = D_{k-1} + H_{k-1}q_{k-1} + L_{k-1}q_{k-1}^2 \tag{A6}$$

All the coefficients in equations (A1) to (A6) can be found by backward induction beginning with the final period n. For any even integer k and any given value of q_{k-1}, we have

$$\begin{aligned} G_k(q_{k-1}) = \sup_r \big[&\pi_2(q_{k-1}, r) \\ &+ \delta\pi_1(q_{k-1}, r) + \rho G_{k+1}(r)\big] \end{aligned} \tag{A7}$$

The value of r that yields the supremum on the right side of equation (A7) is r_k^*. The value of $F_k(q_{k-1})$ can then be determined by the relation

$$\begin{aligned} F_k(q_{k-1}) = \ &\pi_1(q_{k-1}, r_k^*) \\ &+ \gamma\pi_2(q_{k-1}, r_k^*) + \rho F_{k+1}(r_k^*) \end{aligned} \tag{A8}$$

For any odd integer j, similar relations exist for determining the values of $F_j(r_{j-1})$, q_j^*, and $G_j(r_{j-1})$.

Since the values of $G_{n+1}(r_n)$ and $F_{n+1}(r_n)$ must be 0, the backward induction begins with the boundary values

$$A_n = B_n = C_n = D_n = H_n = L_n = 0 \tag{A9}$$

For any even integer k, the following rela-

tions can be derived from equations (A4) to (A8):

$$\text{(A10)}\qquad t_k = \frac{1+\rho H_k}{2(1-\rho L_k)}$$

$$u_k = -\frac{1+\delta}{2(1-\rho L_k)}$$

$$A_{k-1} = \rho D_k - \frac{(1+\delta)t_k^2}{2u_k}$$

$$B_{k-1} = \delta - (1+\delta)t_k$$

$$C_{k-1} = -\left(\delta + \frac{(1+\delta)u_k}{2}\right)$$

$$D_{k-1} = \rho A_k + (\gamma + \rho B_k)t_k + (\rho C_k - \gamma)t_k^2$$

$$H_{k-1} = 1 - (1+\gamma)t_k + (\rho B_k + \gamma)u_k + 2(\rho C_k - \gamma)t_k u_k$$

$$L_{k-1} = (\rho C_k - \gamma)u_k^2 - (1+\gamma)u_k - 1$$

Similarly, for any odd integer j, a completely analogous set of eight equations can be derived by replacing k by j everywhere in (A10) and interchanging γ and δ wherever they occur in (A10).

If the backward induction process is continued indefinitely and the values of t_k and u_k are calculated for successively smaller even integers k, it is found that these values converge to certain limiting values, which we shall denote by T_2 and U_2. Similarly, the values of t_j and u_j for odd integers j will converge to certain limiting values T_1 and U_1. In other words, the reaction functions (A1) and (A4) converge to certain limiting reaction functions of the form

$$\text{(A11)}\qquad q_j^* = T_1 + U_1 r_{j-1}$$

$$\text{(A12)}\qquad r_k^* = T_2 + U_2 q_{k-1}$$

By replacing t_k, u_k, B_{k-1}, B_k, C_{k-1}, C_k, H_{k-1}, and L_{k-1} by their limiting values in (A10), and replacing t_j, u_j, B_{j-1}, B_j, C_{j-1}, C_j, H_{j-1}, and L_{j-1} by their limiting values in the analogous equations for odd integers j, it is found that the limiting values T_1, U_1, T_2, and U_2 must satisfy the four equations shown as (A13). If $\gamma=\delta$, then $T_1=T_2$ and $U_1=U_2$, and it can be found from (A13) that the common value of T and common value of U must satisfy the relations shown as (A14).

Now suppose that $\rho=1$, so that there is no discounting. When $\gamma=\delta$ it follows from equations (7) to (10) that the limiting reaction functions given by (A11) and (A12) will be mutually optimal if, and only if,

$$\text{(A15)}\qquad T = \frac{(1-U)(1+\delta U)}{3+\delta+(1+3\delta)U}$$

$$\text{(A13)}\qquad \delta = \frac{(1+\rho)(1+\rho T_1 - 2T_2) - \rho(2+\rho+\rho U_1U_2)(T_1+T_2U_1)}{\rho[2+\rho+2(1+\rho)U_1+\rho U_1U_2](T_1+T_2U_1) - \rho(1+\rho)(T_1+U_1)}$$

$$\delta = -\frac{(1+\rho U_1U_2)^2 + 2(1+\rho)U_2}{(1+\rho U_1U_2)^2 + 2\rho(1+\rho)U_1^2U_2}$$

$$\gamma = \frac{(1+\rho)(1+\rho T_2 - 2T_1) - \rho(2+\rho+\rho U_1U_2)(T_2+T_1U_2)}{\rho[2+\rho+2(1+\rho)U_2+\rho U_1U_2](T_2+T_1U_2) - \rho(1+\rho)(T_2+U_2)}$$

$$\gamma = -\frac{(1+\rho U_1U_2)^2 + 2(1+\rho)U_1}{(1+\rho U_1U_2)^2 + 2\rho(1+\rho)U_1U_2^2}$$

$$\text{(A14)}\qquad \rho^2(1+\delta)U^4 + 2\rho(1+\rho)\delta U^3 + 2\rho(1+\delta)U^2 + 2(1+\rho)U + 1 + \delta = 0,$$

$$T = -\frac{(1+\rho)(1+\rho\delta U)U}{1+\delta-\rho(1+\delta)U-(\rho^2+3\rho^2\delta+2\rho\delta)U^2-\rho^2(1+\delta)U^3}$$

It can now be verified that if T and U satisfy the relations (A14), then they also satisfy (A15). Hence, the limiting reaction functions (A11) and (A12) are mutually optimal.

An analytic proof that the limiting reaction functions (A11) and (A12) are mutually optimal when $\rho=1$ and $\gamma \neq \delta$ has not been obtained. However, these limiting reaction functions have been computed numerically by carrying out the backward induction process for a wide range of values of γ and δ and it has been verified that in every case the limiting reaction functions are indeed mutually optimal.

Tables 1 and 2 and this Appendix contain a number of interesting and often surprising results which have been obtained from the simple linear model and are worth summarizing.

1. If $\rho=1$, then for any given values of γ and δ, the limiting reaction functions will be mutually optimal.

2. When $\gamma=\delta$ and $\rho=1$, the profits of both firms at the equilibrium point are larger, the larger is the common value of γ and δ. As this common value of γ and δ approaches the value 1, the equilibrium point converges to the Cournot equilibrium point and the profits converge to the Cournot equilibrium profits. However, the reaction functions do not converge to the Cournot reaction functions. They converge to a mutually optimal pair of reaction functions intersecting at the Cournot equilibrium point. It should have been emphasized that the limiting relations (A13) and (A14) have been derived under the assumption that γ and δ are strictly less than 1. These relations are not valid for $\gamma=\delta=1$ because the backward induction process does not lead to a unique equilibrium for these values.

3. If $\gamma<\delta$, then the profits of firm 1 at the equilibrium point will be greater than the profits of firm 2. As firm 1 increases γ up to the level of δ, the profits of firm 1 do not necessarily monotonically decrease. They may continue to increase until γ gets close to δ, and then start to decrease. For example, when $\gamma=.2$ and $\delta=.9$, the profits of firm 1 are greater, and the profits of firm 2 are smaller than they are when $\gamma=0$ and $\delta=.9$.

4. For any fixed value of $\rho (0 \leq \rho < 1)$ if $\gamma=\delta$, then as their common value approaches the value 1, the equilibrium point converges to the collusive solution. Hence, no matter how close to 1 the value of ρ is, as long as $\rho<1$, the profits of both firms converge to the pure cooperative profits as γ and δ converge to 1.

REFERENCES

R. G. D. Allen, *Mathematical Analysis for Economists*, New York 1938.

R. M. Cyert and M. H. DeGroot, (1970a) "Multiperiod Decision Models With Alternating Choice as a Solution to the Duopoly Problem," *Quart. J. Econ.*, Aug. 1970, *84*, 410–29.

——— and ———, (1970b) "Bayesian Analysis and Duopoly Theory," *J. Polit. Econ.*, Sept./Oct. 1970, *78*, 1168–84.

——— and ———, "Interfirm Learning and the Kinked Demand Curve," *J. Econ. Theor.*, Sept. 1971, *3*, 272–87.

F. T. Dolbear et al., "Collusion in Oligopoly: An Experiment on the Effect of Numbers and Information," *Quart. J. Econ.*, May 1968, *82*, 240–59.

———, "Collusion in the Prisoner's Dilemma: Number of Strategies," *J. Conflict Resolution*, June 1969, *13*, 252–61.

J. W. Friedman, "Reaction Function and the Theory of Duopoly," *Rev. Econ. Stud.*, July 1968, *35*, 257–72.

———, "The Theory of Oligopoly," unpublished doctoral dissertation, Yale Univ. 1962.

R. D. Luce and H. Raiffa, *Games and Decisions*, New York 1957.

P. Samuelson, *Foundations of Economic Analysis*, Cambridge 1955.

G. J. Stigler, "Notes on the Theory of Duopoly," *J. Polit. Econ.*, Aug. 1940, *48*, 521–41.

Alternative Programs for Income Redistribution: The NIT and the NWT

By EDGAR K. BROWNING*

Economists have devoted considerable attention to the analysis of the advantages and disadvantages of the negative income tax (*NIT*) as a mechanism for income redistribution. Surprisingly little attention has been given to any significantly different alternative to the *NIT*—most analyses simply compare one variant of the *NIT* to another. But recently several writers have suggested that a form of wage subsidy offers significant advantages over the *NIT*.[1] This paper is an attempt to provide a more complete comparision of these two forms of transfer programs with an emphasis on the aggregate labor market consequences. Needless to say, no attempt is made to consider all repercussions of these policies. Instead, emphasis is focused on labor supply effects and related conditions for economic efficiency.

Section I is an analysis of the effects of implementing a *NIT* when no other transfer programs are present. Section II is a similar analysis for a form of wage subsidy which will be referred to as the negative wage tax, or *NWT*. Section III uses the analysis of the preceding sections to provide a comparison of the *NIT* and *NWT* on two counts: their incidence and welfare cost.

* Assistant professor of economics, University of Virginia. I am grateful to Roland N. McKean, John H. Moore, Edgar O. Olsen, and Roger P. Sherman for comments on an earlier draft. This paper draws on materials from chapters 2, 4, and 7 of my doctoral dissertation. I am especially indebted to Wallace E. Oates, my dissertation supervisor, for his encouragement and assistance.

[1] See Jonathan Kesselman (1969, 1971), Richard Zeckhauser, and Zeckhauser and Peter Schuck.

I. The Negative Income Tax

The *NIT* is the name given to a cash transfer program where the transfer varies inversely with the household's income: the lower the income, the greater the transfer. The mechanics of such a program are probably familiar, but will be summarized briefly in order to show the similarity to the *NWT*. The transfer received by a household, then, is

$$T = r(B - Y) \tag{1}$$

where T is the transfer payment, r is the marginal tax rate (sometimes called the negative tax rate), B is the level of income at which the transfer is zero, which is called the break-even level, and Y is the income of the household. This equation gives the transfer for any household with income below the break-even level.

A *NIT* program is described by its three policy variables: r, B, and the income guarantee, which is simply the transfer that a household with a zero income receives. But only two of the policy variables are independent, since if B and r are set, then equation (1) specifies the transfer where Y = zero (the income guarantee).[2]

In analyzing such a program it will be assumed initially that market wage rates do not change. Figure 1 illustrates the

[2] For a discussion of the relationships among the policy variables, see Christopher Green.

conventional indifference curve analysis of the effects of the *NIT* on work effort and income. *AP* is the budget constraint relating income and leisure in the absence of the *NIT*, with a slope of (−) the wage rate. Equilibrium occurs at *E*, with income of *EE′* and work effort of *PE′*. When the *NIT* is implemented, the constraint becomes *AKM*, where *K* is the break-even income corresponding to *B* in equation (1), and *MP* is the income guarantee. The disposable income of the household will be higher than its earned income for any level of earned income between zero and *B*. The *KM* portion of the constraint is flatter than the original constraint which indicates that the net marginal wage rate of the household is lower under the *NIT*. This is a result of the marginal tax rate of the *NIT*, since the slope of *KM* is equal to $(1-r)$ times the market wage rate. Thus, the higher is r the lower is the net marginal wage rate.

This change in the budget constraint has two implications of importance to our analysis.[3] The first is that the work effort of the household receiving the *NIT* transfer will be reduced. The effect on work effort can be divided into an income effect and a substitution effect. The income effect results from the transfer raising income at each level between zero and the break-even level, and part of this increment will be taken as increased leisure if leisure is a normal good. The substitution effect results from the marginal tax rate which lowers the net marginal wage rate, thus causing a substitution of leisure for income. Assuming leisure to be a normal good, both the income and substitution effects of the program tend to produce a reduction in work effort. The new equilibrium is illustrated at *F* in Figure 1, with

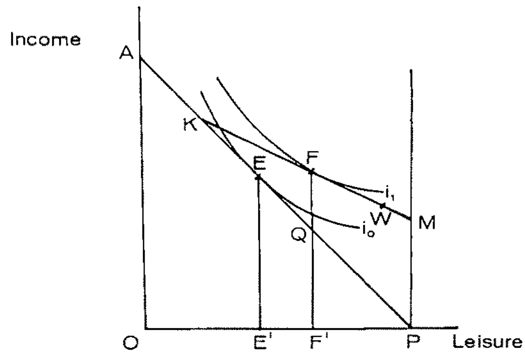


FIGURE 1

work effort equal to *PF′*, earned income of *QF′*, and a transfer of *FQ*. Note that the income of the household has not been increased by the full amount of the transfer because of the reduction in work effort.

That a *NIT* tends to produce a reduction in work effort is well known, but a second implication of the analysis has received much less attention: it is possible that work effort will be reduced sufficiently to reduce the income (money income, not real income, of course) of the household. Thus, an equilibrium may occur at *W* in Figure 1 because the income and substitution effects on income are in opposing directions and the outcome depends on the size of the opposing influences. If the substitution effect is larger then, a *NIT* will actually result in a reduction in the income of the transfer recipient.

This is a somewhat distressing implication of the theory and it thus may be worthwhile to assess its likelihood. The effect on income depends on the size of the substitution effect relative to the income effect: the larger the substitution effect, the more likely is a reduction in income. The substitution effect is a consequence of the tax rate in the program, and since this rate is (normally) the same for all households with income between zero and the break-even level, we can say, roughly, that the stimulus to substitute away from in-

[3] Depending on how the income tax is coordinated with the *NIT*, these implications may also hold for some households with income above the break-even level.

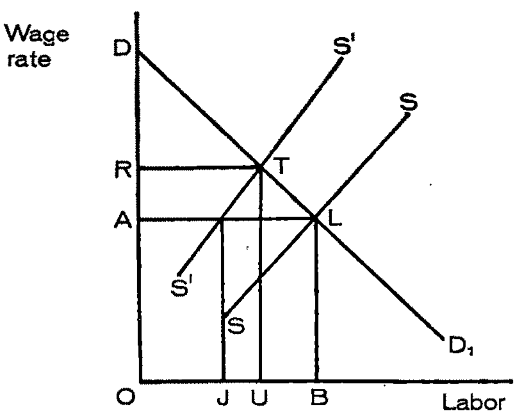


FIGURE 2

come is the same. But the income effect depends on the size of the transfer, and this is smaller the closer is income to the break-even level. Therefore the substitution effect is *relatively* more important the higher is income, and the income effect is *relatively* more important the lower is income, so the likelihood of a reduction in income is greater the higher is income.[4]

So far we have assumed that market wage rates will remain unchanged. This assumption has allowed us to get an idea of the initial reactions to a *NIT*, but this is not the complete picture because wage rates can be expected to change in response to introduction of the *NIT*. If the market for labor services of the transfer recipients is in equilibrium before the *NIT* is introduced, and the *NIT* results in a reduction in the quantity of labor supplied at the existing market wage rate, then the market will no longer be in equilibrium. The quantity demanded will exceed the quantity supplied and the wage rate will be bid up. Thus, the analysis of Figure 1 is incomplete for the final equilibrium cannot lie on the budget constraint *AKM* since this constraint is based on the lower market wage rate that prevailed before the *NIT* was implemented.

To evaluate the effects of this change in wage rates, we have to analyze the aggregate labor market repercussions. In Figure 2, DD_1 and *SS* are the demand and supply curves in the absence of the *NIT* program. They are drawn for the households whose labor services will be in reduced supply as a result of the *NIT*, which we can think of, roughly, as the transfer recipients. *OA* is the initial market wage rate and *OALB* the earned income of the recipients. Since Figure 2 represents an aggregate labor market, *ODLB* is total national income which is divided between the transfer recipients, *OALB*, and the rest of society, *ADL*.[5]

The *NIT* results in a leftward shift in the supply curve to $S'S'$, where the reduction in the quantity of labor supplied at the initial wage rate, *JB*, is simply the sum of the reductions such as $E'F'$ illustrated in Figure 1. Assuming that the other factors of production are in inelastic supply, the demand curve will not shift. Then the equilibrium is at *T*, with a wage rate of *OR* and quantity of labor *OU*. The total income of the transfer recipients will be *ORTU* plus the transfers received (which are not shown in the diagram).

This has several important implications for an analysis of the *NIT*:

1) As we saw earlier, it was possible for the income of the transfer recipients to fall. Once we take account of the increase in the wage rate, however, the likelihood of this occurring is lessened. To see why, think of the total change in income as the

[4] That this is plausible can be seen by examining two extreme cases. If income is initially at the break-even level (*K* in Figure 1), income will definitely be reduced by the *NIT* since there is only a (Slutsky) substitution effect. On the other hand, if income is initially below the income guarantee, then the *NIT* will definitely increase income.

[5] Identifying the area under the demand curve as total income assumes that total income would be zero if the quantity of transfer recipient labor is zero. If this is not the case, then a constant term must be added to *ODLB* to give total income. In addition, identifying *OALB* as the income of the transfer recipients assumes that they have no nonlabor income.

sum of the change which would occur if the wage rate did not change and the subsequent change when the wage rate increases. The first change (illustrated in Figure 1) may result in a reduction in income, as we saw. But the wage rate increase will always increase income from this level, since an increase in the wage rate is the same as a reduction in the "effort" price of income and, hence, more income will be consumed.[6] This will be true regardless of the slope of $S'S'$ as long as the law of demand holds. The net effect on income may still be negative, but it is less likely due to the increase in the wage rate.

2) The change in the earned income of the transfer recipients resulting from the *NIT* will not be an indication of the disincentive effect. Earned income may rise despite a significant reduction in work effort if the wage rate rises sufficiently. If the demand curve is inelastic, then earned income for the entire group will rise following a reduction in labor supply, and conversely.

3) The taxation necessary to finance the transfer cost of the *NIT* will not measure the actual cost borne by the rest of society. This is because the gross (pretax) income of the rest of society will be reduced, reflecting the lower marginal productivity of other factors of production when the quantity of transfer recipient labor they are combined with in production falls.[7] In Figure 2, this reduction in the gross income of the rest of society is from *ADL* to *DRT*, or *ARTL*. The area *ARTL* represents an implicit transfer of income which occurs within the private sector as a result of changes in factor prices and is as much a cost to the taxpayers as the explicit transfer cost of the program.[8]

II. The Negative Wage Tax

The negative wage tax is a transfer program which assists households by subsidizing the prices at which they sell their labor services. The form discussed here is a type of wage subsidy which is similar to the *NIT*. But instead of the subsidy varying inversely with income, as with the *NIT*, it varies inversely with the wage rate of the household. The relationship among the variables is given by

$$S = s(W_s - w) \tag{2}$$

where S is the subsidy per hour of work, s is the rate of subsidy (a fraction), W_s is the break-even wage rate at which the subsidy is zero, and w is the market wage rate.[9] This equation does not represent the total transfer to the household, since the total transfer will be equal to S times the number of hours worked.

Analysis of this program is quite simple. Assume initially that market wage rates do not change when the *NWT* is introduced. Then the budget constraints of all *NWT* transfer recipients will be changed in exactly the same way as if their market wage rates increased, with larger increases the lower is the initial market wage rate. As is well known, the substitution and income effects of such a change are in opposing directions and the impact on work effort is indeterminate. If the household's labor supply curve is upward sloping work effort increases, and conversely if it is backward bending. Without knowledge of the slope of the supply curve no definite prediction can be made.

[6] See Lionel Robbins.

[7] It is possible that the prices of specific factors which are good substitutes for transfer recipient labor will rise. But, on average, the price of all other factors must fall.

[8] In an earlier article in this *Review* I extended this analysis to show that an experimental simulation of a *NIT* will not produce unbiased estimates of the effects of the program.

[9] The similarity between the *NIT* and the *NWT* is apparent from a comparison of equations (2) and (1).

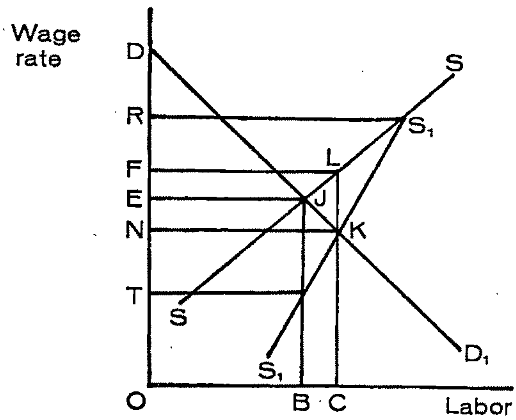


FIGURE 3

Income (the transfer plus earned income) will definitely be increased by the *NWT*. An increase in the wage rate is equivalent to a reduction in the effort price of income, as already mentioned, so more income will be chosen. (In effect, the law of demand places a limit on how backward bending a supply curve can be.)

One other implication should be noted, because of its contrast to the *NIT*. With the *NIT* all households with income below the break-even level will face a net marginal wage rate that is one half their market wage rate (if $r=.5$). This means the substitution effect would be relatively more important compared with the income effect for incomes near the breakeven level. This is not true with the *NWT*: the income and substitution effects will both be smaller near the break-even wage rate because the divergence between the net and market wage rates becomes smaller as we approach the break-even wage rate. For example, if $s=.5$ and $W_s=\$2$, at a market wage rate of \$1 there is a \$.50 difference between the market and net wage rate, whereas at a market wage rate of \$1.90 the difference is only \$.05. Thus, the substitution effect of the *NWT* will tend to be smaller at higher wage rates. As we will see later, this has important implications for assessing the welfare cost of the *NWT*.

Now we will drop the assumption that market wage rates do not change and examine the impact of the *NWT* on the labor market. In Figure 3, DD_1 and SS represent the market demand and supply curves in the absence of the *NWT*: these are the same curves as in Figure 2 if the same households are being subsidized. The equilibrium market wage rate is OE, and the quantity of labor is OB. The gross income of the rest of the community is initially EDJ.

The *NWT* drives a wedge between the market wage rate, which will determine the quantity of labor services employers will wish to employ, and the net wage rate, which will determine the quantity of labor services households will wish to provide. The supply curve with quantity expressed as a function of the market wage rate becomes S_1S_1S. For wages above W_s (OR in Figure 3), the supply curve is unchanged, since there is no subsidy paid for wages above this level. Below this point the vertical difference between the two supply curves is equal to the subsidy per unit. For example, households would be willing to supply OB units of labor at a market wage of OT with the *NWT* in effect, since this implies a net wage rate of OE. With an upward sloping supply curve, therefore, the *NWT* will increase the supply of labor.

This increase in supply results in a new equilibrium at K, with a lower market wage rate (ON) and a higher net wage rate (OF) than previously. If the supply curve is vertical the *NWT* will result in no change in the market rate but a higher net rate, and if it is backward bending the result is a higher market and a higher net wage rate. Only in the perverse case where the demand curve is vertical would the net wage rate received by the recipients fail to increase.[10]

[10] This analysis breaks down if $s=1$. In this case the supply of labor to each *firm* becomes independent of the

The implications of this analysis parallel those discussed with reference to the *NIT*: 1) The total income of the transfer recipients will always be increased by the *NWT* so long as the demand curve is downward sloping. This follows from the fact that the net wage rate will increase regardless of the slope of the supply curve. 2) The change in earned income of the transfer recipients is not an indication of the incentive or disincentive effects of the program since earned income depends on the change in the wage rate as well as the change in hours worked. 3) The transfer cost again does not measure the burden borne by the rest of society. However, for the *NWT*, it is not possible to predict a priori whether the actual burden will be greater or less than the transfer cost, for this depends on the shape of the supply curve. If the supply curve is upward sloping, the actual burden borne by the rest of society ($FLKN-EJKN$) is less than the transfer cost ($FLKN$) by the increase in the gross income of the rest of society ($EJKN$). If the supply curve is backward bending the actual burden is greater than the transfer cost because there will be a reduction in labor supply just as with the *NIT*.

III. Comparison of *NIT* and *NWT*

The analysis in the previous sections gives some useful insights into the workings of two alternative welfare programs, but it is lacking in several respects. In each case the analysis was designed to show the effects of the programs on the obviously unrealistic assumption that no other welfare programs exist. Moreover, the analysis also neglects a number of considerations which are relevant in evaluating a transfer program. For instance, the effects on family size and stability, criminal activity, voting behavior, investment in human capital, and occupational choice may be of considerable importance in evaluating the costs and benefits of a transfer program. However, in a comparison of the *NIT* with the *NWT* at the same level of redistribution, these effects may be kept the same, and therefore not crucial to the analysis. Many of these side effects are undoubtedly related more closely to the income effects of the redistribution than to its particular form, so if we compare programs of equal magnitude we can assume that these side effects will not differ substantially and concentrate on those factors where differences will exist, for example, on labor market effects.

One objection to a comparison of the *NIT* and *NWT* is that the two programs cannot be made comparable since a *NWT* will provide no assistance to those who do not work; whereas under the *NIT* these households will receive the income guarantee. This objection is quite valid and the fact that the *NWT* affects only those who work represents a real drawback. But this drawback is not decisive, for it is possible to use a companion program with the *NWT* to deal with these households. Such a companion program could take several forms, but for our purposes we will simply assume that it provides the same amount of assistance to nonworking households as would the *NIT*. This means that these households will be in similar circumstances under the *NWT* and *NIT*, and we can simply ignore them and concentrate on working households in the analysis.

A. Differential Incidence

Now we will compare the distributional effects of a *NIT* and *NWT* by applying to transfer programs Richard Musgrave's concept of "differential incidence," where-

market wage rate, since the net wage rate would then be W_e regardless of the market wage rate. This implies that the market wage rate would fall to zero and the taxpayers would be bearing all the labor costs. To avoid this it is necessary to maintain adequate differentials in net wage rates for different market wage rates, and this means that s should be significantly below one.

by one tax is conceptually substituted for another and the consequences traced out.[11] The *NIT* and *NWT* must be of equivalent magnitudes for this comparison. But care must be taken in specifying what determines equivalent magnitudes. The most obvious criterion, equal transfer cost, is not appropriate. As we have seen, the transfer cost does not generally measure the burden on the rest of society of either the *NIT* or the *NWT*, so that programs of equal transfer cost may impose unequal burdens on the taxpayers.

Unfortunately, in dealing with a tax plus a transfer program, the definition of equivalence is somewhat arbitrary: one may define equivalence from the standpoint of either the taxpayers or the transfer recipients. Consequently, we will examine both alternatives. Initially, equivalence will be defined as programs which imply the same money income for the transfer recipients. (Ideally, this should be done for equal real income, but this is impossible without a precise knowledge of utility functions and an ability to apply different policy variables to each recipient. Hence, we must content ourselves with equating money income.) And, secondly, equivalence will be defined as programs which imply the same real burden for the rest of society (the taxpayers).

Neither definition is fully acceptable when the analysis is applied to the group case, as ours will be, since the analysis may gloss over distributional effects within each group. For instance, an equal money income for the entire group of transfer recipients under each program may be accompanied with a very different distribution of that income among the recipients. But the seriousness of this is mitigated by the possibility of using graduated marginal tax rates or rates of subsidy, and thereby allowing some control over the distribution of a given money income under either program. Accordingly, we will ignore this complication.

Initially, we will compare programs which result in equal money income for the transfer recipients. The supply and demand curves for the labor services of the recipients in the absence of any welfare program (or in the presence of any welfare program which will remain in effect with both the *NIT* and *NWT*) are shown in Figure 4. Consider the *NWT* plan which shifts the supply curve to S_1S_1. Assume that a *NIT* program which results in the same money income for the transfer recipients shifts the supply curve to S_2S_2. There is no way, of course, to show in the diagram that this *NIT* produces exactly the same money income as the *NWT*. But for our purposes this is not important, since we need only know (as we will see) that the supply curve with the *NIT* lies to the left of the supply curve of the comparable *NWT*. And this will always be the case.[12]

Therefore, when a *NIT* and *NWT* result in the same money income, the quantity of labor supplied will be greater under the *NWT*. The importance of this is that it implies that total national income will be greater under the *NWT* even with a

[11] See Musgrave, ch. 10.

[12] For an upward sloping supply curve this is obvious, since the two programs shift the supply curve in opposite directions. With a backward bending supply curve both programs shift the supply curve to the left, but the shift is greater for the *NIT*. To see why, imagine that the leftward shift is equal for the *NIT* and *NWT*, producing the same market wage rate under both programs. This cannot be an equilibrium, for income and leisure are identical under both programs, yet the net wage rates differ. The same income and leisure cannot represent an equilibrium under both programs, for while there is no difference in the income effects at this point, there remains a substitution effect favoring more work (less leisure) under the program with the higher net wage rate. The *NWT* implies the higher net wage rate, since the net wage under it will exceed the market wage rate, while the net (marginal) wage rate under the *NIT* is always below the market wage rate. This means that an equilibrium with equal income under the two programs must imply greater leisure under the *NIT*.

backward bending supply curve, so the following analysis will remain valid whatever the shape of the supply curve.

Now let us determine under which of the programs there is the greater burden on the rest of society. This can be determined by taking the following measurements from the diagram.

Total national income:

1. Under $NWT = DKCO$
2. Under $NIT = DHAO$

Total income of the transfer recipients:

3. Under $NWT = TKCO +$ transfer cost
4. Under $NIT = FHAO +$ transfer cost

Total money income of the rest of society (which will also be an indication of total real income since we have assumed inelastic supplies of its factors of production) will be:

5. Under $NWT = DKCO - (TKCO +$ transfer cost)
6. Under $NIT = DHAO - (FHAO +$ transfer cost).

This shows that the income of the rest of society is simply the difference between total national income and the income of the transfer recipients. Recall that the transfer programs have been designed so that the income of the transfer recipients (the bracketed terms in the above relationships) are equal, so we can subtract the sixth relationship from the fifth to get:

7. $DKCO - DHAO = HKCA$

$HKCA$ is the amount by which the real income of the rest of society is greater under the NWT than under the NIT. It is simply the difference in national income under the two policies.

Thus, in raising the income of the transfer recipients to any desired level, the real burden on the rest of society is lower under a NWT than under a NIT. Despite the somewhat lengthy derivation above, this is intuitively obvious once it is recognized that national income is greater under the NWT, for the income of the transfer recipients is identical in both cases, and the remainder of total income is the income of everyone else in society.

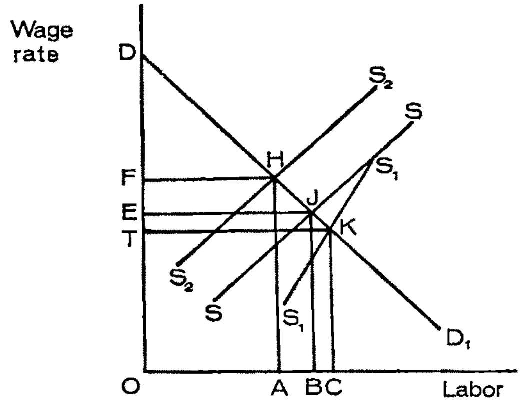


FIGURE 4

It must be noted, however, that equal money income under the two programs does not imply equal real income for the transfer recipients because the quantity of leisure will differ for the two programs. In fact, since leisure will always be greater under the NIT, the transfer recipients (ignoring possible distributional effects within the group) will be better off under the NIT.

Alternatively, we can compare the NIT and NWT when they are designed to produce the same burden on the rest of society, i.e., where the net income of the taxpayers is the same. The same procedure can be used to show that, since national income will still be greater under the NWT, the income of the transfer recipients will be greater under the NWT. Thus, at equal real cost to the rest of society, the NWT raises the income of the transfer recipients by a greater amount. In this case it is not possible to generalize concerning the welfare of the transfer recipients, for while they have greater income under the NWT they have less leisure.[13]

[13] These same results can be derived for an individual using indifference curve analysis, but this approach suffers from the assumption that the market wage rate will be the same under both programs.

B. Welfare Cost

In the absence of market imperfections, any form of non-lump sum transfer payment results in a welfare cost, or excess burden. For the *NIT* the welfare cost reflects the fact that the marginal tax rate drives a wedge between the market wage rate and the net marginal wage rate, leading the recipient to consume too much leisure; if he received the same payment in the form of a lump sum transfer, he would consume more income and less leisure and would be better off. There is also a welfare cost associated with the *NWT*, but here the recipient is led to consume too little leisure and too much income relative to the distortionless lump sum transfer. This difference is due to the fact that the *NWT* involves negative marginal tax rates which result in the net wage rate being greater than the market wage rate, whereas the opposite is true for the *NIT*.

Since both forms of transfer programs produce a welfare cost, it might appear that nothing a priori can be said about their relative sizes. But this is not true. There are good theoretical reasons for believing that the welfare cost of a *NWT* will be smaller than that of a comparable *NIT*. To show this, we will draw on the work of Arnold Harberger concerning the measurement of welfare cost. Harberger shows that the welfare cost of an income tax can be expressed as:

$$\text{(3)} \qquad \text{Welfare cost} = (1/2)em^2wL$$

where e is the elasticity of labor supply (expressing the substitution effect alone),[14] m is the marginal tax rate on income, w is the market wage rate, and L is the quantity of labor supplied.[15] The total welfare cost of the income tax is simply this expression summed over all households.

While this formula was derived by Harberger to measure the welfare cost of an income tax, it can also be used to measure the welfare cost of a transfer program if the transfer payment varies with the household's income. When the transfer varies with income there is an implicit marginal tax rate which drives a wedge between market and net wage rates, and this ensures that the formula is applicable. It makes no difference if the implicit marginal tax rate on income is negative, as it is with the *NWT*.

Now we can give three reasons why a *NWT* is likely to produce a lower welfare cost than a comparable *NIT*.[16]

1. Assume that the marginal tax rate of the *NIT* is .5 and the elasticity of labor supply is the same for all recipients. Then the welfare cost as a percentage of gross earnings is constant for all recipients: it is $(1/2)em^2wL/wL$, or $(1/2)em^2$. Now consider a comparable *NWT* with a rate of subsidy of .5. The welfare cost as a percentage of gross earnings is also $(1/2)em^2$, but for the *NWT* this percentage *falls* as the wage rate of the household is closer to the break-even wage rate. This is because the implicit marginal tax rate on income is not given by the rate of subsidy, but instead is a negative function of the market wage rate.

For example, consider a *NIT* with $B=\$6000$ and $r=.5$, and a *NWT* with $W_s=\$3$ and $s=.5$. At a market wage rate of \$1.50, the negative tax rate (m) under the *NWT* is $-.5$, but at higher wage rates

[14] It is important to recognize that there will be a welfare cost associated with an income tax even if the supply curve as normally drawn, including both income and substitution effects, is perfectly inelastic. The welfare cost reflects the substitution effect of the tax; hence, e in (3) will always be positive.

[15] The derivation of this formula depends on the assumption that distortions other than those arising from the income tax are negligible.

[16] The welfare cost associated with the taxation necessary to finance the programs will be ignored on the assumption that it will be the same for both programs.

it becomes progressively smaller. For instance, at a market wage rate of \$2, m equals $-.25$ since the subsidy here is \$.50 per hour. Thus the *NWT* will involve a lower welfare cost for all households with wage rates above \$1.50 than will the *NIT*, and a higher welfare cost for those with lower wage rates. (This simply reflects the fact that the distortion between market and net wage rates is smaller the higher is the wage rate under the *NWT*, whereas the distortion is the same (in percentage terms) at all wage rates for the *NIT*.) Since most recipients will be in the former category and since the base (ΣwL) is almost certain to be larger here, the total welfare cost is likely to be smaller under the *NWT*.

2. A second factor contributing to a lower welfare cost for the *NWT* is the existence of other taxes and transfers which impose positive marginal tax rates on the recipients of *NIT* or *NWT* transfers. For example, state and local income taxes, general sales taxes, the social security tax, and many state welfare programs all involve positive marginal tax rates.

The existence of these other programs with the same type of welfare cost as the income tax (reflecting their positive marginal tax rates) has two relevant implications. The first is that the additional welfare cost of the *NIT* (one cannot now speak of *the* welfare cost of the *NIT*) will be higher than in the absence of these other programs. The second is that the additional welfare cost of the *NWT* will be lower than in the absence of these programs, and may be negative (a welfare gain).

The welfare cost of a *NIT* as a percentage of gross earnings is $(1/2)em^2$, or is proportional to the square of the marginal tax rate. If $r=.5$, the welfare cost is proportional to .25 if there are no other programs imposing a marginal tax rate on income. But since there are such programs, assume for purposes of illustration that the effective marginal rate (roughly, the sum of the marginal rate of each program)[17] of other programs on income is .2. Then there is a welfare cost prior to implementation of the *NIT* which is proportional to .04. With the *NIT* in effect, the effective marginal rate will be $.7(=.5+.2)$ and the *total* welfare cost will be proportional to $(.7)^2$, or .49. The *additional* welfare cost of the *NIT* will be proportional to $.45(=.49-.04)$. Note that this is substantially higher than the .25 figure we arrive at when we ignore the existence of these other programs.

Thus, the existence of other programs with positive marginal tax rates implies that the additional welfare cost attributable to the *NIT* is larger than it would be in the absence of these programs.

Exactly the opposite is true for the *NWT*: the additional welfare cost is lower when programs with positive marginal tax rates on income exist. The reason is that the effective rate will be the sum of the negative rate of the *NWT* and the positive rate of the other programs: the negative rate of the *NWT* tends to offset the positive rate of the other programs. Or, to put this another way, the distortion of the *NWT* is in the opposite direction from the distortion of the other programs, and this means that the *NWT* will tend to offset some of the distortion introduced by the other programs. For example, at a wage rate where the negative tax rate of the *NWT* is $-.2$, then the total welfare cost will be proportional to $(-.2+.2)^2$, or will be zero since the negative rate of the *NWT* exactly offsets the positive rate of

[17] The effective rate may not be exactly equal to the sum of the marginal rates of each program because the bases of the different taxes may differ. For instance, the effective rate due to the federal income tax and a state income tax is less than the sum because state tax liability is deductible in computing federal tax liability.

the other programs yielding an effective rate of zero. The additional welfare cost of the *NWT* is then proportional to $-.04$, or represents a welfare gain equal to the initial welfare cost of the other programs.

To get an idea of the importance of this, consider again the *NIT* and *NWT* programs discussed earlier (with $B=\$6000$, $r=.5$; $W_s=\$3$, $s=.5$). If the effective tax rate of the other programs is .2, then at any wage rate exceeding \$1.06 the additional welfare cost of the *NWT* will be less than that of the *NIT*. At \$1.06, both programs imply an effective net rate of .7 (in absolute value); for the *NIT* this is $(.5+.2)$ and for the *NWT* $(-.9+.2)$. At higher wage rates the effective rate under the *NWT* is lower. In fact, at wage rates exceeding \$1.67 (where the net effective rate under the *NWT* is $-.2$) there will be a welfare gain as a result of introducing the *NWT*, for the effective rate (absolute value) in the \$1.67–3.00 range will actually be lowered below the preexisting rate of the other programs.

Thus, the *NWT* tends to offset some of the distortions already existing in the system. By contrast, the *NIT* will aggravate these distortions still further.

3. So far the analysis has relied on an implicit assumption which many economists would question, i.e., that a lump sum transfer (with a zero marginal tax rate) is an efficient device (involves a zero welfare cost). This would be the case if there are no externalities associated with the responses of the recipients to the transfer payments. But if there are externalities of certain kinds, then lump sum transfers (which provide the benchmark for a zero welfare cost in (3)) will not be efficient.

Recently, Richard Zeckhauser has argued that externalities may be important in determining the optimal structure of income transfer programs. In effect, he suggests that the money income of the poor is an external economy to the taxpayers, while they do not care (or care less intensely) about the welfare or leisure of the poor.[18] Zeckhauser shows that if this type of interdependence exists then Pareto optimality (i.e., absence of a welfare cost) in a transfer program requires that it have a negative marginal tax rate on income. While his analysis is somewhat involved, the key idea is quite simple: at equal cost to themselves taxpayers prefer (at least up to some point) a transfer program which results in more money income and less leisure for the recipients. At equal cost, a program with a negative tax rate (for example, the *NWT*) will produce more money income and less leisure than a program with a positive rate (for example, the *NIT*) or a zero rate (for example, lump sum transfers). Hence, a program with negative rates will be preferred by the taxpayers. In addition, Zeckhauser's more rigorous formulation demonstrates that the transfer recipient as well as the taxpayer may benefit in moving from a program with a positive rate to a program with a negative rate.

If this assumption is valid, then for a program to involve a zero welfare cost it must use a negative tax rate on income.[19] This implication strengthens our conclusions regarding the relative welfare costs of a *NIT* and a *NWT*. For in this case Harberger's formula must be reinter-

[18] The assumption of this type of interdependence is probably less implausible than most such assumptions. As evidence, it should be noted that virtually all of the literature defines poverty in terms of money income alone, rather than as money income plus the value of leisure time.

[19] Presumably, there is an optimal structure of negative rates rather than a single optimal rate for all the poor. It seems plausible that the optimal rate for a household with an income of \$2000 will be larger (in absolute value) than for a household with an income of \$4000. This would simply reflect a greater concern on the part of the taxpayers to raise the incomes of the poorest households. If this is true, then a *NWT* policy is an appropriate remedy, for the negative rates will tend to be higher for poorer households.

preted: the marginal tax rate (m) in the formula must measure the divergence of the actual marginal rate from the optimal negative rate and not from zero. This means that any *NIT* will result in an even greater additional welfare cost, since the difference between its positive rate and the optimal negative rate will be larger than the difference between its positive rate and zero. On the other hand, a *NWT* will appear even better in comparison with the *NIT*.

These three reasons taken together make a convincing case that a *NWT* will produce a lower welfare cost than a *NIT*. While there is certainly no presumption that any real world *NWT* would be optimal, it will at least tend to alter the tax rate structure in the proper direction, whereas the *NIT* will tend to exacerbate the distortions already present.

One other matter deserves comment. For a *NWT* to function properly it is important that the minimum wage law be repealed. It does little good to subsidize low wage rates heavily if the law prohibits employment at these wage rates. After all, a *NWT* allows the attainment of the objectives of the minimum wage law, which is something that cannot be said for the law itself. Many economists feel that repeal is long overdue in any event, but it is worth emphasizing that it will be all the more important if a *NWT* is implemented.

REFERENCES

E. K. Browning, (1971a) "Incentive and Disincentive Experimentation for Income Maintenance Policy Purposes: Note," *Amer. Econ. Rev.*, Sept. 1971, *61*, 709–12.

———, (1971b) "Income Redistribution and the Negative Income Tax: A Theoretical Analysis," unpublished doctoral dissertation, Princeton Univ. 1971.

C. Green, *Negative Taxes and the Poverty Problem*, Washington 1967.

A. C. Harberger, "Taxation, Resource Allocation, and Welfare," in *The Role of Direct and Indirect Taxes in the Federal Revenue System*, Princeton 1964.

J. Kesselman, "Conditional Subsidies in Income Maintenance," *Western Econ. J.*, Mar. 1971, *9*, 1–20.

———, "Labor Supply Effects of Income, Income-Work, and Wage Subsidies," *J. Hum. Resources*, summer 1969, *5*, 275–92.

R. Musgrave, *The Theory of Public Finance*, New York 1959.

L. Robbins, "On the Elasticity of Demand for Income in Terms of Effort," *Economica*, June 1930, *10*, 123–29.

R. Zeckhauser, "Optimal Income Transfers," *Amer. Econ. Rev.*, June 1971, *61*, 324–34.

——— **and P. Schuck**, "An Alternative to the Nixon Income Maintenance Plan," *Publ. Interest*, spring 1970, *19*, 120–30.

Simulation of the Impact of Social and Economic Institutions on the Size Distribution of Income and Wealth

By FREDERIC L. PRYOR*

In the vast literature on the distribution of income, relatively little systematic attention has been paid to the mutual influence of the distributions of income and wealth on each other, and to the impact of intergenerational grants on the entire process. The purpose of this essay is to utilize a simulation model in which the distributions of income and wealth are analyzed together. The model permits me to explore the influences of such socioeconomic variables as the pattern of intergenerational grants, the rules of inheritance, the patterns of mate selection, differential fertility of various income classes, and the patterns of governmental redistributions of income and wealth. So many simplifying assumptions must be made that the model, as it is presented below, cannot be directly used for determining the current size distribution of income and wealth; nevertheless, the results of the model suggest certain neglected factors, particularly the shape of the intergenerational savings function, that must be taken into account if we are to gain a clearer picture of the causal forces operating in the real world. The most important biases (mostly leading to greater income equality) arising from the simplifying assumptions of the model are also discussed below.

To take all the major long-run factors influencing the size distributions of income and wealth into account is an extremely complicated matter, and we face two alternative research strategies; either we model these factors mathematically, which requires some drastic simplifications in order to keep the equation system solvable,[1] or we take more factors into account by simulating their impact. While the latter procedure does not lead to a completely general solution it does permit us to investigate certain features of economic systems using parameters of particular interest.

To place this model in perspective, it is useful to note that variance of personal income can be derived from three sources; variances in the distribution of labor income, of property income, and of the interaction between these two variables.[2] In this model, variations of labor income are handled by assigning everyone a lifetime income equal to the average lifetime in-

* Professor of economics, Swarthmore College. I wish to thank Howard Pack, J. Roland Pennock, Frank C. Pierson, Zora Pryor, and George Stolnitz for their helpful suggestions on an earlier draft. The research for this essay was financed by the International Development Research Center of Indiana University and was carried out at the Economic Growth Center of Yale University. A considerable amount of difficult work was done by those who programmed this extremely complicated simulation model, and I would especially like to thank the chief programmer, Carol Hopkins. I would also like to express my appreciation to David Forman and Michael Hooven for assistance at particular points.

[1] Such an approach is followed by Joseph Stiglitz and, in a less formal fashion, by James Meade.

[2] The higher the correlation between individual labor and property incomes, the more unequal the distribution of income. (See formula in fn. 11.) The correlation can, of course, be negative, a situation that apparently arose in mandarin China where receivers of property income made a special point of avoiding manual work, even if it meant a lifetime of poverty. (See Hsiao-Tung Fei.)

come times a random variable with a mean of unity and a specified standard deviation. Variations in property income stem only from differential holdings of wealth; returns per unit of wealth are assumed equal; and, finally, no correlation is assumed between individual labor and property incomes. Thus the critical factor for a change in the distribution of income is a change in the distribution of wealth which, in turn, is greatly influenced by the different socioeconomic variables specified below.

I. The Basic Model[3]

The simulation model starts with 100 unmarried people with an arbitrary initial distribution of productive wealth. These people are "put to work" and both lifetime labor and property incomes are generated by means of a production function (a Cobb-Douglas function is used at first but a *CES* function is later tried) with assumptions of full employment and of mean factor payment equal to its marginal product. The total amount of property income is divided among all wealth holders in amounts proportionate to the quantity of wealth held by each; the total amount of labor income is distributed to the entire population by giving each a wage equal to the average wage times a normally distributed variable (which is supposed to represent a differential distribution of abilities such as intelligence or diligence).

The people in the model are then lined up according to income, and marriages are arranged according to one of three different rules: 1) A person can only marry another person next to him on the income scale (this is called the no-choice rule); 2) The chances for a person marrying anyone else are equal (the equal-choice rule); 3) A person can marry anyone but the chances are greater if the two are closer to each other on the income distribution (the limited-choice rule).[4]

The government can then step in and redistribute income (either progressively or regressively). This is followed by the accumulation or disposal of family wealth (positive or negative savings) so that a specified ratio of family wealth to family income is achieved; this intergenerational savings function is discussed in detail below. A simplifying assumption used in the model is that all wealth and net changes in wealth (net savings or dis-savings) are in the form of productive capital which yield property income.

At this point the various families have children according to their income in the following manner: First, the families are divided into three groups according to whether they are among those with the highest family incomes, lowest incomes, or in-between. (The percentage of families falling in each group can be varied.) Then the number of children are specified for each group, for example, the rich can be designated to have more or fewer children than the poor (or vice versa). Polygynous situations can be approximated by specifying many children for the rich and no children for the poor, since in such societies it is well known that only the wealthier can afford to support many wives and the low income men often do not marry.

The parents are then removed from the scene and family wealth is divided among

[3] The model rests on the pioneering work of Guy Orcutt. Other simulation models of income distribution have been made (for example, Hans-Juergen Krupp), but these are considerably different than mine and, moreover, have focused primarily on short-run problems.

[4] The probabilities of one person marrying another is, of course, changed when a couple is married and removed from the pool of eligibles. Therefore, the statements in the text must be considered as just approximately true. For the limited-choice model, the probability of marriage is inversely proportional to the difference in income rank between the two individuals. The calculations are simplified by not designating the sex of the individuals so that marriages between any two individuals may be possible.

the children according to one of three different rules: 1) One child can receive everything (the primogeniture rule);[5] 2) The property can be divided equally among all children (the equal-division rule); 3) The first child can receive half of the wealth and the remainder is divided equally among the rest (the compromise rule). At this point the government can also redistribute wealth (a type of inheritance tax). We now have a group of people with a given distribution of wealth of productive capital whom we put to work; the process is repeated many times to see if a stable distribution of income and wealth is achieved.

With only a few exceptions (discussed below) the processes converge toward an "equilibrium distribution" which, when attained, is maintained for all succeeding generations. In certain cases, however, the process is extremely slow and in order to avoid inordinate computer expenses, the following procedure was adopted. In every case the simulation was carried out twice, once starting from a highly unequal initial distribution of wealth (where the wealthiest 10 percent of individuals share the total societal wealth in equal portions) and once starting from a relatively equal wealth distribution (where the wealthiest 75 percent of individuals share in equal amounts the total wealth). Each simulation is then run for 30 generations (which represents 1000 years if a generation is calculated as $33\frac{1}{3}$ years, or 750 years if a generation is 25 years) and the end results of the simulations starting from different points are averaged; in almost all cases the two estimates were very similar to each other.

[5] I have used an extreme form of primogeniture in which the first child, regardless of sex, obtains the entire family's wealth. This variant of primogeniture leads to the most extreme wealth-holding inequalities and, in addition, is computationally simpler in that sexes do not need to be assigned to particular individuals for the model to work. A more usual case of primogeniture is, of course, when only the eldest son obtains the entire estate.

In the tables I use Gini coefficients as a measure of this inequality of the equilibrium size distribution of lifetime income. Gini coefficients are measured by calculating Lorenz curves and measuring the ratio between the "area of inequality" and the total income triangle (the coefficient ranges from 0.00, which represents total equality to 1.00, which represents total inequality). I have computed four other statistical measures of inequality but space does not permit their inclusion. In certain cases where it is useful to discuss relative speeds of convergence, the period is measured from the starting point to the point when an equilibrium Gini coefficient is achieved (plus or minus a small amount).[6]

Before the numerical results are presented, certain features of the model may perhaps be better understood if we examine in a qualitative fashion the effect of particular variables.

A. The Simplest Patterns

If we take a situation where every family, regardless of income, has two children; where there is no variation in labor income (the random factor is not yet introduced); and where families do not add to, or decrease, their inherited wealth (i.e., family wealth is passed on unchanged regardless of income, a situation occurring when land is the basic source of wealth and is not alienable), the observed patterns of convergence of the distribution of wealth are quite simple and are outlined in Table 1 below.

With primogeniture wealth accumulates eventually into a single hand.[7] The speed of convergence is inversely proportional to the degree to which people choose marriage partners in other income brackets. With

[6] The exact method employed is described in the Appendix.

[7] If the eldest son variant of primogeniture is used, then, of course, this extreme result will not obtain.

TABLE 1—BASIC EQUILIBRIUM PATTERNS OF THE DISTRIBUTION OF WEALTH

	Marriage rules		
Inheritance rules	No-choice rule	Limited-choice rule	Equal-choice rule
Primogeniture	Wealth concentrates to a single owner; fast convergence.	Same as no-choice rule, but convergence is slower.	Same as no-choice rule, but convergence is slower than other primogeniture situations.
Equal division	Wealth distribution remains the same as starting position.	Wealth becomes evenly distributed but convergence is slow.	Wealth becomes evenly distributed; convergence is faster than with limited-choice rule.

equal division of inherited property and with everyone marrying a person next to him on the income distribution, it should be readily apparent that no change will occur in the size distribution of income and wealth, as long as all families continue to have two children. In the other two cases with equal division of inherited property, wealth eventually becomes completely equally distributed.

The compromise rule of inheritance, where the first child received half of the property and the remainder is divided equally among the other children, is the same as the equal-division rule (under the assumption of two-child families).

The speed of conversion to equilibrium depends, of course, on the distribution of wealth at the starting point and at equilibrium. Starting from a highly unequal distribution of income, convergence is achieved in the primogeniture case in three to five generations; starting from the more equal distribution of wealth, convergence takes six to fifteen generations. For the equal-division cases where convergence occurs, the process generally takes somewhat longer. Although convergence speeds in the above examples appear to have little economic meaning, they become important when governmental policy measures are introduced (i.e., redistribution of income and/or wealth) and certain goals of income distribution are set.

B. The Impact of a Random Variable Representing Abilities

If we now introduce a random variable representing differential abilities so that labor incomes are the product of the average wage times a random variable (with a mean of unity), the basic patterns are modified in the following ways:

1) The greater the variation in the random element, the more unequal the equilibrium income distribution in those five cases in Table 1 where convergence is observed. This is because greater extremes in labor income are generated.

2) In situations where people are allowed choice in marriage partners, the convergence process is speeded up where there is an equal division inheritance rule and slowed down where there is primogeniture. This stems from the fact that there is a greater mixing of people of different wealth at the time of marriage (since the marriage rules are based on total income, not wealth alone). Where people marry those next to them on the income scale, a complication arises which is discussed below.

One important methodological point also arises. Introduction of a random element raises difficulties in determining the exact equilibrium distribution of income. In interpreting the various tables presented below, small differences in the coefficients should be overlooked and, al-

though data are presented to three places, variations of several percent or less should usually be neglected.

Throughout the remaining simulations the random factor is set with a standard deviation of 15 percent, which is roughly similar to the variation of I.Q. test scores.

C. The Impact of Differential Fertility Rates

In all cases primogeniture leads to a highly unequal income distribution since, whatever the fertility pattern, only one child receives all the property. Similarly, in all cases where the rich have only one child (which leads automatically to primogeniture) a highly unequal equilibrium income distribution is also generated.

In nonprimogeniture cases, fertility affects the end results considerably. In cases where fertility increases with income, the system should converge relatively quickly to fairly even distributions. (Such a situation allegedly occurred in past eras in oriental despotic societies where the rulers encouraged polygyny so that the rich would have a much higher fertility rate than other classes and, at the same time, forbade primogeniture; the end result was supposed to be an economy with a relatively even distribution of income and no independent bases of wealth with which to challenge royal authority.)[8]

Simulation results of different patterns of fertility among the income classes are presented in Appendix Table 1. For simplicity, in the rest of the paper, I assume that fertility is the same in all income classes unless otherwise specified.

II. Quantitative Results: No Saving; Stationary Population; No Technological Change

We are now ready to begin the quantitative investigation. In Table 2 below, the results for the most simple situation with different marriage and inheritance rules are presented. There is no accumulation of wealth for the society as a whole and each family passes on only that wealth which it inherits.

[8] Such systems are analyzed by K. Wittfogel.

TABLE 2—GINI COEFFICIENTS OF EQUILIBRIUM INCOME DISTRIBUTION ASSUMING DIFFERENT MARRIAGE AND INHERITANCE RULES

	Marriage Rules		
Inheritance Rules	No-Choice	Limited-Choice	Equal-Choice
Primogeniture	.307	.308	.297
Compromise	[a]	.064	.060
Equal Division	[a]	.064	.060

Assumptions: No net family capital formation; no capital or income redistribution; all families have two children; standard deviation of the random element is .15; labor share of national income is 75 percent.

[a] If the system starts from a highly unequal distribution of wealth, the equilibrium distribution of income is equal to its original value. If the system starts from a relatively equal wealth distribution where high ability people with no property might marry low ability people with property (since they would be next to each other on the income scale), then the equilibrium income distribution would be highly equal.

As expected, the inequality of income is higher with primogeniture than with the compromise or equal division inheritance rules (which give the same answers because each family has only two children). As we also expect, the equilibrium income distribution does not seem greatly affected by the marriage rules; in this simple model the major effect of the marriage rules appears on the speed at which equilibrium is achieved. Only in much more complicated models do the marriage rules appear to have much impact on the equilibrium size distribution of income. One puzzling phenomenon appears in the primogeniture case where the other measures of inequality give somewhat different results but this is due most likely to a change in the shape of the distribution of income, with slightly increasing inequality at the high income end and slightly de-

creasing inequality at the middle and lower income levels. This particular result does not appear due to random factors since equilibrium was achieved from all starting points of the simulation and the various results given particular marriage and inheritance rules were very similar.

In the simulation model a wealth redistribution process can be set up so that all inherited wealth is taxed a given percentage and then the total amount of taxed wealth is distributed equally among all individuals. If the tax rate is positive, then such a process is progressive because the least wealthy end up with a net gain in wealth while the most wealthy end up with a net loss. This manner of specifying a capital tax on inherited wealth allows the total amount of privately held wealth to remain constant.

The results of imposing a capital redistribution tax are presented in Table 3 and can be summarized quite easily: The greater the redistribution of wealth, the more equal the equilibrium distribution of income and wealth. This, of course, is not surprising. The equilibrium income distributions are most unequal for the primogeniture cases and most equal when family property is evenly distributed among heirs, a result similar to the previous findings. The marriage rules again have a relatively small impact. It must be noted that with very high redistributions of wealth, the differences in the equilibrium distributions with the various marriage and inheritance rules are relatively small; it appears that after a certain point, the redistribution swamps the effects of other institutions.

Since the redistribution of wealth can be a deliberate tool of governmental policy, the rate of convergence to the equilibrium income distribution is of considerable interest. Several generalizations can be made: First, imposition of a redistribution of wealth greatly increases the speed of convergence and in almost all cases convergence is achieved within five generations (with the major portion of the changes occurring in the first two generations). Second, convergence appears to be faster when the model is started with relatively more unequal wealth distribution

TABLE 3—GINI COEFFICIENTS OF EQUILIBRIUM INCOME DISTRIBUTION WITH DIFFERENT REDISTRIBUTIONS OF WEALTH

Redistribution and inheritance rules	Marriage Rules: No-Choice	Limited-Choice	Equal-Choice
No redistribution of wealth			
Primogeniture	.307	.308	.297
Equal division	[a]	.064	.060
30 percent redistribution of wealth			
Primogeniture	.156	.158	.148
Equal division	.061	.063	.063
60 percent redistribution of wealth			
Primogeniture	.088	.091	.089
Equal division	.062	.062	.064
90 percent redistribution of wealth			
Primogeniture	.061	.066	.064
Equal division	.062	.062	.062

Assumptions: see Table 2
[a] See fn. a, Table 2

than with a relatively more equal wealth distribution. In other words, achieving a more equal distribution of income by means of a redistribution of wealth can be carried out more quickly in those cases where the differences between actual and desired distributions are greatest.

Introducing a redistributive income tax changes, of course, the equilibrium income distribution. In this case wealth is not affected; and since marriages are arranged according to relative income, this part of the system is not affected either. Under the assumptions in this section about no-net intergenerational accumulations of family capital, the speed of convergence toward the equilibrium income distribution should not be affected. However, once we allow net intergenerational accumulation of capital based on family income, then an income redistribution could have a number of effects on total savings, growth, and convergence and provides a much more interesting problem for analysis. Discussion of the effects of a redistributive income tax is therefore deferred until the rigid assumption about the integenerational transfer function is loosened.

III. Quantitative Results: Introduction of an Intergenerational Savings Function

Let us now allow the capital stock to change, while still keeping the total population constant, so that we can isolate some of the effects of different patterns of intergenerational savings and transfers. One of the most important results of the simulation exercise is to show that the shape of this function has crucial importance on the equilibrium income distribution. Before turning to the results, however, several theoretical questions deserve brief examination.

First we must inquire about the nature of the integenerational transfer function. Up to now this has been a *terra incognita* in the economic literature: empirical data with which to derive such a function are unavailable and theorists who insist on deducing propositions about consumer behavior from the neoclassical axioms of rational choice have not been able to say anything about the matter. One reasonable assumption, which is reflected in numerous *obiter dicta* on the subject, is that the amount of wealth passed on by a husband and wife to their children is primarily a function of lifetime family income and it is on this basis that the model is changed below. The shape of the function is still problematical and experiments are carried out with several different formulae.

Second, once we introduce changes in the net capital stock, a problem arises because we must also take into consideration the possibility of multiple stable growth paths;[9] thus the initial conditions of the system and the way in which the intergenerational transfer function are specified become quite important. In order to avoid such complications, care was exercised in designing the intergenerational transfer function so that only a single stable growth path, independent of the initial conditions, would be achieved. A number of tests were also made with different initial conditions (different initial capital stocks) in order to insure the correctness of the specification.

A third problem arises in the choice of the production function, i.e., the function showing the relationship between the capital and labor in the system and total production. All production functions used below exhibit diminishing returns to variations in the capital-labor ratio. This means, among other things, that the ratio of capital to labor asymptotically approaches a limit which represents a "steady-state equilibrium" where depreciation just equals gross savings and net

[9] This problem is analyzed in an abstract but lucid manner by Stiglitz.

capital formation is zero. Such diminishing returns can, however, be offset by technological change so that in "steady-state growth" production, the capital stock and the ratio of capital to labor rise at the same rate of growth. In this particular case (Harrod-neutral growth), labor productivity (the ratio of output to the constant labor force) also rises at the same rate while capital productivity asymptotically approaches a constant.

One last precautionary note must be added. Although the analysis below focuses almost exclusively on income distribution, it must be emphasized that the various types of intergenerational transfer functions, production functions, and redistributions of wealth and income lead to quite different equilibrium levels of production. Although there are a number of propositions in the economic literature linking the inequality of income to the growth rate of production (for example, the more unequal the distribution of income, the greater the aggregate production), the results below show exceptions to such generalizations. For those interested in pursuing the relationship between income distribution, the savings function and growth, some data on steady-state production equilibria are given in Appendix Tables 2, 4, and 6.

A. *The Impact of Linear and Non-Linear Intergenerational Transfer Functions*

Two types of integernerational transfer functions are used in the analysis below. The first is a simple linear function where intergenerational transfers of a particular family are a proportion of lifetime income: $S=Yz$ where S=intergenerational transfers; Y=family lifetime income; and z is the "savings constant" to be specified. If family income is low and inherited wealth is great, the intergenerational transfer (proportionate to income) might be less than the original inherited wealth, i.e., the family has "dipped into capital." A second type of intergenerational transfer function has a kink in it and is thus non-linear; such transfers are a function of income over and above some socially determined "subsistence level," below which no intergenerational transfers are made (i.e., the family doesn't pass on inherited wealth). Since negative intergenerational transfers (i.e., debts) are not permitted (although dipping into capital is allowed), the kink of the savings function occurs at $\overline{Y}$. The simple form selected for this non-linear function is $S=(Y-\overline{Y})z$, $S\geq 0$.

A brief digression is necessary to clear up certain ambiguities of this subsistence level. First, this is not necessarily the biological subsistence level nor does there need to be any explicit societal recognition that such a subsistance level actually exists. Rather, it is merely the income level below which families feel they must spend all of their funds for consumption in order to try to achieve a certain standard of living and as a result, they do not have any wealth left over to pass on to succeeding generations. If per capita income in the society rises and the socially determined subsistence level remains stationary, this subsistence level may eventually become such a small proportion of individual family income that the savings function is, for all intents and purposes, linear, i.e., the performance of the system asympototically approaches that of a system with an intergenerational transfer function of $S=Yz$. On the other hand, the socially determined subsistence level can also rise as society views on an adequate standard of living rise. Since this might be a reasonable approximation of reality, it is useful to tie the subsistence income to the rise in per capita income. For simplicity, I have set the subsistence income always equal to average per capita income which means that when a non-linear intergenerational transfer function is used, only those

TABLE 4—GINI COEFFICIENTS OF EQUILIBRIUM INCOME DISTRIBUTION ASSUMING DIFFERENT INTERGENERATIONAL TRANSFER FUNCTIONS

Transfer functions and inheritance rules	Marriage Rules: No-Choice	Limited-Choice	Equal-Choice
	$S=Yz$		
$z=1.5$			
Primogeniture	.165	.165	.161
Equal division	.069	.067	.060
$z=2.0$			
Primogeniture	.167	.169	.162
Equal division	.066	.066	.063
$z=2.5$			
Primogeniture	.165	.170	.164
Equal division	.065	.063	.062
	$S=(Y-\bar{Y})z$		
$z=2.0$			
Primogeniture	.306	.301	.300
Equal division	.293	.266	.206
$z=2.5$			
Primogeniture	.308	.309	.306
Equal division	.296	.260	.199

Assumptions: All families have two children; standard deviation of random element is .15; labor share of national income is 75 percent; no income or capital redistributions; no technical change; no negative transfers.
Note: S=intergenerational transfers; Y=personal incomes; $\bar{Y}$=average income; z=a constant.

families with incomes above the average pass on wealth to the succeeding generation. Finally, it must be noted that the socially determined subsistence level can rise faster than average income and in this strange case with such a strong "demonstration effect," families might dip into capital (as long as they did not run into debt) until little capital would be left in the society as the subsistence level approaches the highest family incomes.

Experiments were also made with a third form of the intergenerational transfer functions: $S=W+Yz$, where W is the wealth inherited by the mother and father of the family from their parents. This function implies that the wealth inherited by the children is always greater than the wealth inherited by the parents, no matter how low the family income might be because of the low labor incomes received by the parents. Such a situation does not seem very realistic and these experiments were abandoned. The results of simulation experiments using both linear and non-linear intergenerational transfer functions for several different parameters of the savings coefficient (z) are presented in Table 4 above.[10]

The non-linear transfer function leads,

[10] If the capital-output ratio is greater than unity, the average z coefficient must be greater than unity or there must be a multiplicative constant in the production function if the capital stock in the economy is to be maintained. In the results reported in Table 4, z is placed larger than unity. If a multiplicative constant were used in the production function, z could be made less than unity (which, of course, makes more "real life" sense) but the results would be the same.

as one might expect, to a much greater inequality of income than the linear function; for in the former case, only the richer segments of the population pass on wealth to their children and this, in turn, concentrates wealth and property income. With a simple linear transfer function, the height of the savings coefficient (z) does not appear to affect the inequality of the equilibrium income distribution (the differences do not appear statistically significant) although, of course, a higher z leads to a higher steady-state production level. On the other hand, a higher savings coefficient does seem to affect the degree of income inequality when a non-linear function is used, although the direction of the effect depends upon the particular measure of inequality that is chosen and the inheritance rule that is followed.

With the introduction of intergenerational transfer functions based on income, we now have a situation where the marriage rules have a more important impact on the equilibrium income distribution than in the previous section where capital was passed on regardless of income. With primogeniture the effect of the marriage rules is small in almost all cases except with one nonreported inequality measure. On the other hand, with equal division of property, especially with a non-linear transfer function, the inequality of the equilibrium income distribution decreases as the marriage rules change toward equal choice.

Generalizing about the relationsips between equilibrium income distribution, inheritance rules, and transfer functions is more difficult. With a linear intergenerational transfer function, primogeniture leads to a more unequal distribution of personal income than other inheritance rules. With a non-linear transfer function, the results depend upon the measure of inequality chosen since the relationships between different parts of the income distribution are differentially affected. Thus no generalization is possible.

To summarize, with a linear transfer function where families at all income levels pass on wealth to the succeeding generation, the primary influence on the equilibrium distribution of income is the inheritance rules; and marriage rules or the height of the savings constant (z) have little effect. With a non-linear intergenerational transfer function, the equilibrium income distribution is affected by the marriage rules, the inheritance rules, and the height of the savings constant. Generalizations in these latter cases are difficult because the overall inequality and the shape of income distribution curve change simultaneously.

B. The Impact of Income and Capital Redistributions

A redistribution of capital in each generation not only affects the distribution of income but also, in the case of a non-linear intergenerational transfer function, the growth of the system. (This is because the share of income over the subsistence level is a smaller ratio of total income after a redistribution). The effect on the equilibrium distribution of income should increase with the severity of the redistribution and, one might suspect, would greatly affect the equilibrium distribution in the case of a non-linear, rather than a linear, transfer function. Relevant data are presented in Table 5 below which supports these conjectures.

An X percent redistribution of income in each generation should have a greater impact on the equilibrium distribution of income than an X percent redistribution of wealth because in the latter case only one source of income inequality is being changed. This differential impact of an income redistribution should also be greater in the case of a non-linear intergenerational savings function because the effect on the

TABLE 5—GINI COEFFICIENTS OF EQUILIBRIUM INCOME DISTRIBUTION WITH DIFFERENT CAPITAL REDISTRIBUTIONS

Redistribution taxes and inheritance rules	Marriage Rules No-Choice	Limited-Choice	Equal-Choice
	$S=Yz$, $(z=2.0)$		
No redistribution			
Primogeniture	.167	.169	.162
Equal division	.066	.066	.063
30 percent capital redistribution			
Primogeniture	.121	.120	.123
Equal division	.061	.050	.060
60 percent capital redistribution			
Primogeniture	.086	.081	.087
Equal division	.063	.064	.063
90 percent capital redistribution			
Primogeniture	.065	.064	.064
Equal division	.062	.062	.062
	$S=(Y-\overline{Y})z$, $(z=2.0)$		
No redistribution			
Primogeniture	.306	.301	.300
Equal division	.293	.266	.206
30 percent capital redistribution			
Primogeniture	.233	.229	.226
Equal division	.220	.190	.126
60 percent capital redistribution			
Primogeniture	.144	.145	.133
Equal division	.134	.103	.089
90 percent capital redistribution			
Primogeniture	.069	.070	.068
Equal division	.063	.062	.064

Assumptions: See Table 4

share of income over the subsistence level vis-à-vis the total national income would be greater. And finally, the effect of an income distribution should increase with the extent of the redistribution and (using the same arguments employed in the case of the capital redistribution) should affect the equilibrium distribution more in the case of the non-linear than of the linear transfer function. These conjectures receive support in the data presented in Tables 5 and 6 although certain exceptions do arise, especially when other inequality measures are used. Although the underlying reasons for these exceptions are obscure, part of the difficulty may lie in the changing shape of the income distribution curve that accompanies the overall changes in equilibrium inequality.

Again, we observe the fact that convergence to the equilibrium income distribution is very much speeded up with the imposition of either a redistribution of income or wealth tax program. Again, convergence is achieved usually in a few generations although the parameters of the equilibrium are somewhat different than those discussed in Section II.

Since these income or capital taxes may serve as deliberate policy tools by the government to achieve particular income distribution goals, we must turn briefly to the convergence properties of the system. In the case of the linear intergenerational transfer functions, the situation is very similar to the case where there was no net capital formation in each family and convergence occurs very rapidly, usually in

TABLE 6—GINI COEFFICIENTS OF EQUILIBRIUM INCOME DISTRIBUTION WITH DIFFERENT INCOME REDISTRIBUTIONS

Redistribution taxes and inheritance rules	Marriage Rules No-Choice	Limited-Choice	Equal-Choice
	$S=Yz,\quad (z=2.0)$		
No redistribution			
Primogeniture	.167	.169	.162
Equal division	.066	.066	.063
30 percent income redistribution			
Primogeniture	.112	.115	.112
Equal division	.046	.044	.043
60 percent income redistribution			
Primogeniture	.062	.064	.062
Equal division	.024	.026	.024
90 percent income redistribution			
Primogeniture	.015	.016	.015
Equal division	.006	.006	.006
	$S=(Y-\overline{Y})z,\quad (z=2.0)$		
No redistribution			
Primogeniture	.306	.301	.300
Equal division	.293	.266	.206
30 percent income redistribution			
Primogeniture	.216	.215	.212
Equal division	.202	.176	.130
60 percent income redistribution			
Primogeniture	.124	.123	.121
Equal division	.118	.105	.064
90 percent income redistribution			
Primogeniture	.031	.031	.030
Equal division	.029	.026	.019

Assumptions: See Table 4

several generations. In the case of the non-linear intergenerational transfer function, the system approaches the equilibrium distribution equally rapidly; but in the wealth redistributions, the Gini coefficient of income inequality often cycles around the equilibrium value several times, rather than asymptotically approaching it.

Since the greatest changes in income equality brought about by the income or wealth redistribution occur in the first generation and since differences between the actual degree of income inequality and the equilibrium value are quite small after a few generations, one lesson seems clear: other things being equal, redistribution taxes appear an efficient method of changing the degree of income inequality in a nation. (One condition in the *ceteris paribus* clause is that such taxes do not have adverse effects on productivity.) Although radicals may be unwilling to wait a generation for the major effects to take place and prefer instead a revolution to accomplish redistribution aims, the destruction of capital and confusion following such events might lead to a situation where average income would be considerably lower and income not much more equally distributed than if mundane and undramatic redistributional taxes had been used instead.

C. The Impact of the Elasticity of Substitution

Up to now the results are based on a Cobb-Douglas production function which

has an elasticity of substitution of unity. Since we have growth in the system, this assumption may have some impact on the results and a series of simulations were run using the *CES* function that has been explored by Kenneth Arrow et al.[11]

Summarizing the results in a capsule form is extremely difficult for a number of counteracting factors influenced the results: further, the results feature a number of small puzzles which are difficult to explain. For those wishing to explore these matters further, the results are presented in Appendix Tables 3 and 4.

IV. Quantitative Results: Introduction of Certain Dynamic Factors

A. The Impact of Harrod-Neutral Technological Change

If we introduce technological change such that the production arising from a given capital and labor stock is multiplied by an exponentially growing factor, per capita income grows, the capital stock grows (because savings increase), and eventually the system achieves a steady-state growth path. Using a Cobb-Douglas production function, factor income shares remain the same.

The results of the simulations can be easily summarized: the equilibrium income distributions do not seem greatly affected by the introduction of Harrod-neutral technical change. (Indeed, the only noticeable effects occur with the non-linear transfer function and these are relatively small.) Undoubtedly other types of technological change would complicate the results, but exploration of these matters must be left for future research. Results of the experiments with Harrod-neutral change are presented in Appendix Table 5.

B. The Impact of Population Change

If we now add to the analysis population growth which can arise from many different patterns of differential fertility, a large number of cases are open to explore. Simplification can be achieved once we realize that introduction of population change has two major effects: it raises the absolute value of total production; and it allows different rates of growth of the capital-labor ratio to occur through changes in the denominator of the fraction, rather than the numerator. Rather than multiply examples endlessly, it seems most useful to examine only several simple patterns of fertility in order to show how the system works. One financial constraint on this process of analysis must also be mentioned: the greater the number of people in the system, the more expensive the simulation becomes. To obtain the results reported in Table 7, I started the system with only fifty people; introduced a 10 percent population growth (per generation) and ran the system for only twenty-five generations. This led to an eightfold increase in population and, as a result, almost a quadrupling of computer cost. (It must also be noted that in order to limit population growth, "poor" and "rich" families are defined as the 20 percent of families on either end of the income distribution, while the "middle class" is the remaining 60 percent; this is slightly different than the income definitions used in calculating Appendix Table 1 below.) Equilibrium gross national products are presented in Appendix Table 6.

The most surprising result appears where the transfer function is linear: here differential fertility appears to have relatively little impact on the equilibrium distribution of income, a result which is somewhat different from the situation in Appendix Table 1 where no net capital formation takes place. In the case of the nonlinear transfer function, on the other hand, the expected impact of differential fertility can be observed in nonprimogeniture situations, i.e., the equilibrium

[11] See Arrow et al. For the derivation of factor shares I used the simple formulae derived by R. G. D. Allen.

TABLE 7—GINI COEFFICIENTS OF EQUILIBRIUM INCOME DISTRIBUTION WITH POPULATION CHANGE AND DIFFERENT FERTILITY PATTERNS

Number of children: Rich	M.C.	Poor	Marriage Rules: No-Choice	Limited-Choice	Equal-Choice
			$S=Yz,\ (z=2.0)$		
2	2	2			
Primogeniture			.167	.169	.162
Compromise			.066	.066	.063
Equal division			.066	.066	.063
3	2	2			
Promogeniture			.082	.178	.175
Compromise			.071	.068	.071
Equal division			.065	.069	.066
2	2	3			
Primogeniture			.182	.176	.175
Compromise			.075	.076	.073
Equal division			.074	.073	.072
			$S=(Y-\overline{Y})z,\ (z=2.0)$		
2	2	2			
Primogeniture			.306	.301	.300
Compromise			.293	.266	.206
Equal division			.293	.266	.206
3	2	2			
Primogeniture			.314	.313	.309
Compromise			.273	.253	.169
Equal division			.230	.185	.127
2	2	3			
Primogeniture			.312	.313	.310
Compromise			.309	.296	.194
Equal division			.306	.294	.219

Assumptions: Standard deviation of random element is .15; Cobb-Douglas production function with labor share of national income as 75 percent; no technological change; no income or capital redistribution; no negative transfers.

distribution appears more equal, the greater the number of children of the rich vis-à-vis other groups in the population.

It should also be noted that population growth gives rise to a somewhat more unequal equilibrium distribution of income than with no population growth and it seems likely that this effect would be greater if population growth were higher. This may be tied up with the results that with population growth, the equilibrium per capita income and the equilibrium capital-labor ratio are somewhat lower which, as a result, means that returns per unit of capital are higher and returns per unit of labor are lower. The exact interaction of these various factors is, however, complex.[12]

[12] Some insight can be gained into these matters by starting from the well-known formulas for the separation of the components of variance;

$$\mathrm{Var}(Y)=\mathrm{Var}(P)+\mathrm{Var}(W)+2\ \mathrm{Cov}(P,W)$$

and $\mathrm{Var}(P)\approx \bar{\imath}^2\ \mathrm{Var}\ (k)+\bar{k}^2\ \mathrm{Var}(i)+2\bar{\imath}\bar{k}\mathrm{Cov}(i,k)$

(over)

V. Application of the Model

Given the assumptions of the model, the numerical results obtained in the simulations cannot be directly applied to available data on the distribution of income. Certain biases resulting from the assumptions require special attention.

First, the model assumes that labor income of an individual is not positively correlated with his parents' income. Since it is generally believed that there is, indeed, a positive correlation, this means that the results presented above have a bias toward equality. In future simulations this fact could be built in the model either directly (a procedure which would require considerably more memory capacity of a computer than the program used in this study) or by using a different type of intergenerational savings function in which a fraction of the parents' income would be considered human capital transmitted to the children.

Second, the model assumes that labor income and property income are not positively correlated. Although data have not been published in a useable form empirically to investigate such matters, I strongly suspect that in the United States there is a positive correlation. Certainly those occupying important positions to which high labor incomes accrue are in a better position to invest their money in investments yielding high returns; further, the existence of great wealth often permits people to obtain positions yielding high labor incomes. If this correlation between labor and property income is positive, then the empirical results obtained with the simulation model show a greater income equality than actually exists. Repairing this fault in the model would not be difficult: labor income could be made a function of the random variable plus a given fraction of wealth.

Third, for technical reasons labor incomes in the model were bounded by limits of .5 and 1.5 of the average income, and since capital accumulation is a function of total income, certain limits are placed on the amount of capital accumulation that one individual can carry out. Since this does not permit the existence of a Henry Ford, a J. Paul Getty, or a John D. Rockefeller, who manage to accumulate enormous sums within a single lifetime, the results of the simulation model show a bias toward equality. This might be repaired in the model by designating one person in each generation who is destined to strike it rich at the expense of everyone else (who is "taxed" for this purpose).

Fourth, the distribution of income is calculated from labor and property income before family accumulation (or disaccumulation) takes place. This procedure omits a source making for greater income inequality, namely the income accruing to owners of recently accumulated wealth. In the model presented above, this should not make very much difference; but in more complicated models, this factor must be taken into consideration. On a more general level, the simulation model is based on the assumption that capital accumulation for the entire society occurs through the net addition to inherited productive capital by various families in the system. An alternative method for achiev-

---

where Y = total personal income; P = property income; W = work income, i = return on wealth; k = wealth; $\text{Var}(Y)$ = variance of Y; $\text{Cov}(i, k)$ = covariance between i and k; and a bar over a letter indicates an average value. In the case discussed in the text, $\bar{i}^2$ increases, $\text{Var}(k)$ remains the same, the $\bar{k}^2\text{Var}(i)$ factor still remains zero (since there is no variation in return per unit of capital) and the covariance term remains roughly the same. Thus the overall variance in property income rises and this, in turn, leads to an increase in the variance of overall personal income. Certain other puzzling phenomena remain, particularly in the data for the non-linear savings function; these are due in part to the fact that unlike most other simulations in this essay, equilibrium was achieved extremely slowly and often by the twenty-fifth generation (the cutoff point), this equilibrium point had not been reached.

ing economic growth that does not involve inheritance occurs when claims on productive capital are accumulated through part of a person's lifetime and then are converted into consumption by the end of the person's life. If generations overlap and if the maximum accumulated wealth is greater for each succeeding generation, then societal capital accumulation could occur without any inheritances.[13] Although such extreme a situation does not seem very likely, such a process may be occurring in part and the equilibrium income results of the simulation model would have a further bias toward equality. This factor could be incorporated into the model, but considerably more memory space in the computer would be required.

Fifth, in the simulation model factor incomes are distributed according to the marginal productivity theory. This is not, however, a crucial aspect of the model since by substituting different coefficients in the production function one can easily change the share of income received for the services of labor and capital. Although I have not considered any interactions between relative factor shares and some of the other parameters of the system (on the grounds that no convincing evidence of such a relationship has been found), such an effect could easily be incorporated into the simulation.[14]

Sixth, the model is allowed to run to equilibrium which, in many cases, takes twenty or so generations. Much more useful for policy purposes would be examination of situations where, starting with the current size distribution of income and wealth, the model were allowed to run only for several generations. The purpose of letting equilibrium be achieved is to give some general results about the direction of change that may be of use to analysts facing many different size distributions of income and wealth.

Seventh and finally, our knowledge about the actual parameters of the system that would influence the equilibrium size distribution of income and wealth is quite limited. Most importantly, we have no statistical idea about the shape of the intergenerational savings function which proves such a critical factor in determining the final equilibrium positions. We do not know the relative importance of more complicated inheritance arrangements where wealth is passed two generations away through particular types of trust arrangements. Our quantitative notions about marriage patterns and inheritance rules may be greater, but such matters still need considerable quantitative analysis before parameters can be derived for use in the model. Except for some imaginative work by Robert Summers, little work has been done on calculating lifetime size distributions of income.[15] Thus, even if the simulation model were more sophisticated, we would not have the requisite knowledge of the proper parametrs for running the model for predictive purposes.

VI. Some Speculations

Although the simple simulation model has some obvious shortcomings and must be viewed as a starting point for more sophisticated models, the results obtained point toward a number of factors neglected by economists interested in the size distribution of income. Moreover, the results permit several different answers to some puzzling questions regarding changes in the degree of inequality of the size distribution of income over time.

Among Western non-Marxist economists there seems to be some agreement that the distributions of income and wealth

[13] Such a situation is explored by James Tobin.

[14] I would like to thank Steven Resnick for his ideas on this theme.

[15] Lifetime income distributions in the United States are estimated by Summers. I would like to thank him for sending me a copy of this extremely useful study.

of nations become increasingly unequal during the early stages of the development process but that after the development process is well underway, this distributional tendency is reversed and the income distribution becomes increasingly more equal. Certain empirical evidence supports this proposition. For instance, in the United States estimates of the distribution of wealth show an increasing inequality throughout the ninteenth century, reaching a high point at the turn of the century. Since then, however, these data show that the inequality of wealth holding has declined so that the Gini coefficient of wealth inequality was roughly the same in 1962 as in 1860.[16] In this century the inequality of personal wealth holdings appears to have decreased in the United Kingdom as well.[17] Finally, income inequality has decreased in a great many developed nations in the last 30 to 60 years and there appears to be an inverse relationship in the West between the level of development and the Gini coefficient of income inequality on both a time-series and a cross-section basis.[18] A number of theoretical arguments concerning the causes of such shifts in the distribution of income have been offered[19] and in this digression I would like to add one additional explanation based on the simulation results above.

In a highly underdeveloped nation the major source of wealth is land; inheritances consist primarily of intergenerational transfers of a fixed amount of land; and the equilibrium distribution of income in such a case is described in Section II. As industrialization begins and accumulated industrial capital becomes an important source of wealth, it seems likely that the intergenerational transfer function approaches the non-linear form described in Section III. This is because the biological level of subsistence is still a substantial proportion of average income and it is unlikely that people with relatively low incomes could pass on a very significant proportion of their lifetime income to their heirs. As per capita income rises, this biological subsistence income becomes an increasingly smaller share of average incomes and it seems likely that the income level below which no intergenerational transfers take place does not rise as fast as average income. If so, then the non-linear intergenerational transfer function asymptotically approaches the linear case.

In such a situation, three stages in the distribution of wealth and income can be distinguished: a stagnant stage in which the distribution of wealth and income remain relatively constant; the initial stages of industrialization in which the distribution of wealth and income become increasingly more unequal (under the impact of a non-linear intergenerational savings function); and a later stage of industrialization in which the distribution of wealth and income become more equal when the intergenerational transfer function becomes more linear.

Since we know very little about intergenerational transfers at any stage of development, this scenario of development

[16] For relevant data see Lee Soltow, Robert Lampman, and James Smith.

[17] Data are presented by Lampman, p. 214.

[18] The four most extensive recent international comparisons of the size distribution of income are by Simon Kuznets, Irving Kravis, Harold Lydall (who only covers labor income), and Richard Weisskoff.

[19] Kuznets (1955) focuses on the shift from rural to urban areas as the most important casual factor. (This model is investigated more throughly on a theoretical level by Henri Theil, and on an empirical level by Weisskoff.) Stiglitz bases an explanation for the same phenomenon on the relationship in the process of economic growth of the starting point to the equilibrium production level. R. Albert Berry focuses on unemployment and changes in particular market imperfections. Others have focused on more politcal factors such as the increase of political mobilization accompanying economic development that leads to greater progressive redistribution of income and wealth by the government after a particular point of development.

must remain speculative; nevertheless, it does provide a focus for future empirical research.

VII. Two General Conclusions

A great many different cases have been discussed, and these are but a small fraction of the possible cases that can be generated by the model. Nevertheless, two general conclusions can be drawn:

First, for a general theory of the size distribution of income, we must take into account the influence of the size distribution of wealth; and this means we must bring into the analysis a number of important social and economic variables such as the pattern of intergenerational transfers of income and wealth, the rules of inheritance in the society, and differential patterns of fertility of income classes. It is impossible to generalize about long-run changes in the size distribution of income and wealth in capitalism without specifying many more variables than economists have usually done. Blanket predictions about increasing concentration of income and wealth, for example, the orthodox Marxist analysis of such problems, implicitly make too many vital assumptions to be of much use.

Second, the simulation model presented in this paper provides a starting point for such a broader type of analysis of the distribution of income and wealth. In order to bring the analysis closer to actual situations in particular countries, much more empirical and theoretical work needs to be done. On the theoretical side, we need to consider many more complications than those presented in this essay; on the empirical side we need to have a much clearer picture of the critical parameters. The model does, however, point to one extremely important factor—the shape of the intergenerational savings function—which has been neglected by previous analysts and which I hope to have demonstrated is critical in predicting changes in the size distribution of income and wealth.

Appendix

Determination of the Speed of Convergence to an Equilibrium Income Distribution

Due to the influence of random factors in the inheritance-marriage simulation program, the generated income distribution does not completely converge to a single income distribution, but rather to a band of income distributions around the equilibrium. A number of curve-fitting methods were attempted in order to derive the equilibrium distribution but these proved unsatisfactory and the following alternative method was adopted. This is a modification of a method suggested by Richard N. Cooper, to whom I would like to express my thanks.

First, an unweighted average and standard deviation of the Gini coefficient of income equality were calculated for the last five generations in the thirty-generation simulation. The sixth to last Gini coefficient was then tested to see if it fell within the .95 confidence limit of the calculated average. If this was the case, then the unweighted average and standard deviation were recalculated to include this datum and the next Gini coefficient was examined in a like manner. This process was stopped when the examined coefficient did not meet the test; the number of generations was then determined, and average coefficient for the other indicators of inequality (for example, the standard deviation of the logarithms of income, the share of income accounted for by the top 10 percent, etc.) were recorded.

Then, with the calculated standard deviation of the Gini coefficient from the above process, I started from the first generation to see at what generation the Gini coefficient had a significant chance of belonging to the calculated equilibrium. When this point was reached, the generation number was recorded. The conversion point was considered to lie between this point and the earliest generation to be included in the calculation outlined in the first step.

Such a procedure is based on the assumption that the income distribution converges by and large within thirty generations. Whether such a convergence occurred at all was determined by visual inspection of the entire series of calculated Gini coefficients.

APPENDIX TABLE 1—GINI COEFFICIENTS OF EQUILIBRIUM INCOME DISTRIBUTIONS ASSUMING DIFFERENTIAL FERTILITY RATES[c]

Number of children[a]			Marriage Rules		
Rich	M.C.	Poor	No-Choice	Limited-Choice	Equal-Choice
2	2	2			
Primogeniture			.307	.308	.297
Compromise			b	.064	.060
Equal division			b	.064	.060
3	2	1			
Primogeniture			.309	.309	.306
Compromise			.075	.074	.076
Equal division			.074	.069	.074
1	2	3			
Primogeniture			.308	.308	.298
Compromise			.303	.304	.301
Equal division			.308	.300	.295
1	3	1			
Primogeniture			.310	.306	.300
Compromise			.306	.305	.295
Equal division			.308	.302	.295
3	1	3			
Primogeniture			.310	.306	.303
Compromise			.168	.160	.135
Equal division			.168	.158	.130

Assumptions: No net family capital formation; no capital or income redistributions; standard deviation of random element is .15; labor share of national income is 75 percent.

[a] The poor are those 25 percent of families with the lowest income; the rich are those 25 percent of families with the highest incomes.

[b] If the system starts from a highly unequal distribution of wealth, the equilibrium distribution of income is equal to its original value. If the system starts from a relatively equal wealth distribution where high ability people with no property might marry low ability with property (since they would be next to each other on the income scale), then the equilibrium income distribution would be highly equal.

[c] A Gini coefficient of zero represents total equality; a coefficient of unity represents total inequality.

APPENDIX TABLE 2—EQUILIBRIUM GROSS NATIONAL PRODUCTS WITH COBB-DOUGLAS PRODUCTION FUNCTIONS AND DIFFERENTIAL SAVINGS FUNCTIONS[a]

	Marriage rules					
	No-choice		Limited choice		Equal choice	
	Inheritance rules					
	Primogeniture	Equal division	Primogeniture	Equal division	Primogeniture	Equal division
$S=Yz$: any income or capital redistribution						
$z=1.5$	114	114	114	114	114	114
$z=2.0$	126	126	126	126	126	126
$z=2.5$	136	136	136	136	136	136
$S=(Y-\overline{Y})z$						
No redistributions						
$z=2.0$	79	78	79	75	79	71
$z=2.5$	85	84	85	79	85	73
$z=2.0$, income redistributions of R percent						
$R=0.0\%$	79	78	79	75	79	71
$R=30.0$	70	69	70	65	70	57
$R=60.0$	58	57	58	55	58	47
$R=90.0$	37	36	37	34	37	31
$z=2.0$, capital redistributions of R percent						
$R=0.0\%$	79	78	79	75	79	71
$R=30.0$	70	70	70	66	70	58
$R=60.0$	58	58	58	54	58	52
$R=90.0$	47	48	49	48	47	47

Assumptions: All familes have two children; standard deviation of random element is .15; labor share of national income is 75 percent; no negative saving; no technical change.

Notes: S=personal savings; Y=personal income; $\overline{Y}$=average income; z=a constant.

[a] For the equilibrium *GNP* using the non-linear saving function, production at the 30th generation was used as the equilibrium value.

Appendix Table 3—Gini Coefficients of Equilibrium Income Distribution with Different Production Functions

Production functions, transfer and inheritance rules	Marriage Rules: No-Choice	Limited-Choice	Equal-Choice
	$S=Yz,\ (z=2.0)$		
CES, $s=1.5$			
Primogeniture	.208	.208	.200
Equal division	.061	.060	.059
Cobb-Douglas			
Primogeniture	.167	.169	.162
Equal division	.066	.066	.063
CES, $s=0.5$			
Primogeniture	.102	.101	.102
Equal division	.079	.076	.074
	$S=(Y-\overline{Y})z,\ (z=2.0)$		
CES, $s=1.5$			
Primogeniture	.234	.233	.237
Equal division	.220	.188	.117
Cobb-Douglas			
Primogeniture	.306	.301	.300
Equal division	.293	.266	.206
CES, $s=0.5$			
Primogeniture	.404	.402	.405
Equal division	.400	.377	.327

Assumptions: All familes have two children; standard deviation of random element is .15; no technological change; no income or capital redistribution; no negative transfers.

Notes: S=intergenerational transfers; Y=personal income; $\overline{Y}$=average income; z=constant; L=labor force; K=capital stock; P=total production; b=a constant; s=elasticity of substitution.

Production functions: Cobb-Douglas: $P=L^{.75}K^{.25}$

CES: $P=[.75L^{-b}+.25K^{-b}]^{-1/b}$, where $b=(1/s)-1$

Appendix Table 4—Equilibrium Gross National Products with Different Production Functions[a]

Production functions	No-choice: Primogeniture	No-choice: Equal division	Limited-choice: Primogeniture	Limited-choice: Equal division	Equal-choice: Primogeniture	Equal-choice: Equal division
			$S=Yz,\ (z=2.0)$			
CES, $s=1.5$	131	131	131	131	131	131
Cobb-Douglas	126	126	126	126	126	126
CES, $s=0.5$	117	117	117	117	117	117
			$S=(Y-\overline{Y})z,\ (z=2.0)$			
CES, $s=1.5$	75	74	75	72	75	67
Cobb-Douglas	79	78	79	75	79	71
CES, $s=0.5$	86	85	86	82	86	73

(Column heads: Marriage rules — No-choice, Limited-choice, Equal-choice; Inheritance rules — Primogeniture, Equal division.)

Assumptions: All families have two children; standard deviation of random element is .15; no negative savings; no technical change; no capital or income redistributions.

Notes: S=personal savings; Y=personal income; $\overline{Y}$=average income; z=a constant; s=elasticity of substitution.

[a] See fn. a, Appendix Table 2.

APPENDIX TABLE 5—GINI COEFFICIENTS OF EQUILIBRIUM INCOME DISTRIBUTION WITH HARROD-NEUTRAL TECHNICAL CHANGE

Transfer functions and inheritance rules	Marriage Rules: No-Choice	Limited-Choice	Equal-Choice
$S=Yz$, $(z=2.0)$		No technical change	
Primogeniture	.167	.169	.162
Equal division	.066	.066	.063
		Technical change = 10% per generation	
Primogeniture	.165	.165	.161
Equal division	.069	.067	.060
		Technical change = 20% per generation	
Primogeniture	.165	.165	.161
Equal division	.069	.067	.060
$S=(Y-\overline{Y})z$, $(z=2.0)$		No technical change	
Primogeniture	.306	.301	.300
Equal division	.293	.266	.206
		Technical change = 10% per generation	
Primogeniture	.308	.305	.309
Equal division	.300	.253	.190
		Technical change = 20% per generation	
Primogeniture	.308	.305	.309
Equal division	.300	.253	.191

Assumptions: All families have two children; standard deviation of random element is .15; labor share of national income is 75 percent; no income or capital redistribution; no negative savings.

Notes: S = intergenerational transfers; Y = personal income; $\overline{Y}$ = average income; z = a constant.

APPENDIX TABLE 6—EQUILIBRIUM GROSS NATIONAL PRODUCTS WITH DIFFERENT FERTILITY RATES AND POPULATION GROWTH[a]

Fertility Patterns: Number of children — Rich	M.C.	Poor	No-choice: Primogeniture	No-choice: Compromise	No-choice: Equal division	Limited-choice: Primogeniture	Limited-choice: Compromise	Limited-choice: Equal division	Equal-choice: Primogeniture	Equal-choice: Compromise	Equal-choice: Equal division
						$S=Yz$, $(z=2.0)$					
3	2	2	535	535	535	535	535	535	535	535	535
2	2	2	535	535	535	535	535	535	535	535	535
						$S=(Y-\overline{Y})z$, $(z=2.0)$					
3	2	2	337	314	298	336	309	281	336	271	252
2	2	2	337	335	335	336	331	328	336	282	288

Assumptions: Standard deviation of random element is .15; Cobb-Douglas production function with labor share of national income is 75 percent; no technological change; no income or capital redistribution; no negative savings; rich and poor are top and bottom 20 percent of income distribution, respectively.

Basic parameters: 50 familes, 25 generation

Abbreviations: S = personal savings; Y = personal income; $\overline{Y}$ = average income; z = a constant

[a] For the equilibrium GNP production at the 25th generation was used as the equilibrium value.

REFERENCES

R. G. D. Allen, *Macro-Economic Theory*, New York 1967.

K. Arrow et al., "Capital-Labor Substitution and Economic Efficiency," *Rev. Econ. Statist.*, Aug. 1961, *43*, 225–50.

R. A. Berry, "Income and Wealth Distribution in the Development Process and their Relationships to Output Growth," Yale Economic Growth Center disc. pap. 89, 1970.

H.T. Fei, "Peasantry and Gentry: An Interpretation of Chinese Social Structure and Its Changes," *Amer. J. Sociol.*, July 1946, *52*, 1-17.

R. E. Gallman, "Trends in the Size Distribution of Wealth in the Nineteenth Century: Some Speculations," in L. Soltow, ed, *Six Papers on the Size Distribution of Income and Wealth*, New York 1969.

I. B. Kravis, *The Structure of Income: Some Quantitative Essays*, Philadelphia 1962.

H.-J. Krupp, *Theorie der personellen Einkommensverteilung*, West Berlin 1968.

S. Kuznets, "Economic Growth and Income Inequality," *Amer. Econ. Rev.*, Mar. 1955, *45*, 1–28.

———, "Quantitative Aspects of the Economic Growth of Nations," Part VIII, "Distribution of Income by Size," *Econ. Develop. Cult. Change, 11*, Part 2, Jan. 1962.

R. J. Lampman, *The Share of Top Wealth Holders in National Wealth, 1922–56*, Princeton 1962.

H. Lydall, *The Structure of Earnings*, Oxford 1968.

J. E. Meade, *Efficiency, Equality and the Ownership of Property*, London 1964.

G. H. Orcutt et al., *Microanalysis of Socioeconomic Systems: A Simulation Study*, New York 1961.

J. Smith, "Income and Wealth of Top Wealth Holders in the United States, 1958," unpublished doctoral dissertation, Univ. Oklahoma.

L. Soltow, *Six Papers on the Size Distribution of Income and Wealth*, New York 1969.

J. E. Stiglitz, "The Distribution of Income and Wealth among Individuals," *Econometrica*, July 1969, *37*, 382–97.

R. Summers, *An Econometric Investigation of the Size Distribution of Lifetime Average Annual Income*, tech. rep. no. 31, dept. of economics, Stanford Univ. 1956.

H. Theil, "Migration and its Effect on Per Capita Income Inequality: Maxwell's Demon on Ellis Island," *Economics and Information Theory*, Chicago 1967, 114–20.

J. Tobin, "Life Cycle Saving and Balanced Growth," in W. Fellner et al., *Ten Economic Studies in the Tradition of Irving Fisher*, New York 1967.

R. Weisskoff, "Income Distribution and Economic Growth: An International Comparison," unpublished doctoral dissertation, Harvard Univ. 1969.

K. Wittfogel, *Oriental Despotism: A Comparative Study of Total Power*, New Haven 1957.

The Production Lag

By JOHN A. CARLSON*

Ascertaining the nature of the transmission and diffusion of economic adjustments and of the speed with which they occur offers a host of perplexing problems. We shall focus here primarily on one such problem: finding the production lag or period of time from which production is begun on an item until it emerges from the productive process as a finished good. The emphasis will be on the use of accounting data to determine the length of this lag and its relation to other data reported about manufacturing industries.

The sequel is as follows. Section I develops a production-period formula; Section II explores the relationship between the production period and whether an industry produces primarily to order or to stock; Section III adds considerations of the average delivery period; Section IV investigates a conjecture by Charles Holt and Franco Modigliani about variations in the production period; Section V, with distributed lead estimates, attempts to obtain more information than the industry data appear willing to reveal; and Section VI is a summary of the main results.

I. A Production-Period Formula

Discrete counterparts to the continuous relationships presented here may be found in Holt and Modigliani, pp. 35–40. For notation, let

x = the production period

$q(t-u+x)$ = rate of production of finished goods at time $t-u+x$

$w(t, t-u)$ = work-in-process at time t of items whose production began at time $t-u$

$m(y)$ = rate of materials added per unit of goods in production for a length of time y

$v(y)$ = rate of value-added per unit of goods in production for a length of time y

W_t = total stock of work-in-process at time t

The gross value of production of finished goods equals total materials used and total value-added. Thus, for one unit of output:

$$\int_0^x [m(y) + v(y)]dy = 1 \tag{1}$$

We can also write

$$\begin{aligned} &w(t, t-u) = q(t-u+x) \int_0^u [m(y)+v(y)]dy \\ &\text{for } 0 \leq u \leq x; \text{ and} \\ &w(t, t-u) = 0, \text{ otherwise} \end{aligned} \tag{2}$$

The integral can be thought to represent the percentage completion after time u in production. This percentage multiplied by $q(t-u+x)$, the rate of finished production that will occur at time $t-u+x$, gives $w(t, t-u)$, the corresponding level of in-process stock at time t. Then, integrating over all levels of work-in-process yields the total stock of work-in-process at time t:

$$\begin{aligned} W_t &= \int_0^x w(t, t-u)du \\ &= \int_0^x q(t-u+x) \int_0^u [m(y)+v(y)]dy\, du \end{aligned} \tag{3}$$

Work-in-process inventories thus depend on future output, the rate at which materials and value-added are inserted into the

* Professor of economics, Purdue University. Without implicating Moses Abramovitz, Gerald Childs, Robert A. Meyer, and Victor Zarnowitz in any shortcomings of this paper, I am grateful to them for their comments on an earlier draft.

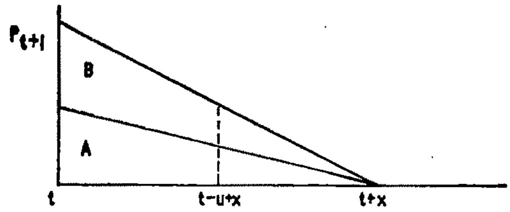


FIGURE 1a. ASSUMPTIONS FOR EQUATION (5) ILLUSTRATED

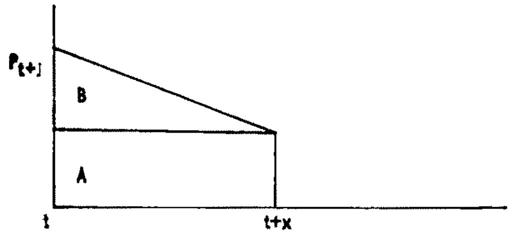


FIGURE 1b. ALL MATERIALS ADDED AT BEGINNING

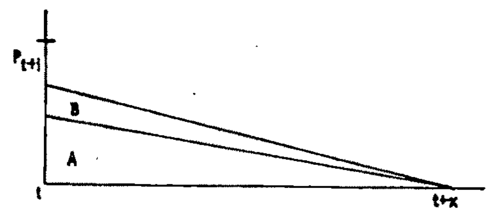


FIGURE 1c. MARKUP AT END

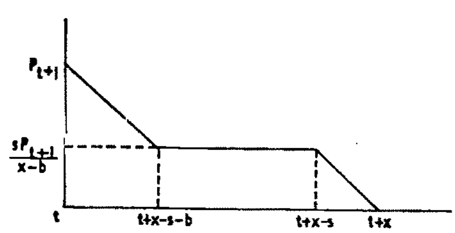


FIGURE 1d. CASE WITH TWO STAGES OF PRODUCTION

FIGURE 1. GRAPHICAL RELATIONSHIPS BETWEEN THE PROJECTED COMPLETION DATES (FROM t TO $t+x$), PRODUCTION PERIOD x, MATERIALS CONSUMED (A), VALUE-ADDED IN PRODUCTION (B), AND TOTAL WORK-IN-PROGRESS INVENTORIES AT TIME t.

production process, and the production period. Data are available on work-in-process inventories and on production. If we knew or could assume something about the rate at which goods-in-process are written up in value, we could infer a value for the production period x.

In the empirical estimates below, a month will be taken as one unit of time. Production of finished goods for the month ending at time $t+1$ will be:

$$P_{t+1} = \int_t^{t+1} q(z)\, dz \tag{4}$$

Consider a stationary situation in which $q(t-u+x)$ is a constant q. In that case, monthly production $P_{t+1}=q$. If we also assume that $m(y)$ and $v(y)$ are constants, then from (1), $m+v=1/x$. Making this substitution in (3) and assuming production is constant yields the following when (3) is integrated out: $W_t=qx/2$. Substituting P_{t+1} for q and solving for x, we have

$$x = 2W_t/P_{t+1} \tag{5}$$

Equation (5) was anticipated by Moses Abramovitz, whose interest in the timing of cyclical turning points of inventory components led him to this technique for estimating the production period.[1]

Figure 1a depicts our assumptions graphically. Future production is assumed to remain constant at level P_{t+1}. The total area of the two triangles A and B represents W_t, total stock of work-in-process

[1] This concept of the production period differs in two respects from that associated with the Austrian measure of the roundaboutness of a productive process as explained by Bohm-Bawerk, pp. 78–91. First, we take fixed capital as given and evaluate the time within manufacturing plants of transforming inputs into finished goods; Bohm-Bawerk includes in the figure for the period of production the time taken to produce the fixed capital as well. If one were interested and had enough information about the productive relationships, the Austrian figure could be pieced together from the figures obtained by the approach taken here. Second, Bohm-Bawerk takes a mean of the time from various labor inputs until final output, whereas we assume a steady accretion of value-added from the beginning of a particular productive process and compute the overall time from start to finish. Geometrically, our production period corresponds to the base of a triangle and his to approximately half the base (ignoring the time to produce fixed capital).

inventories at time t, and the height of the line at $t-u+x$ represents $w(t, t-u)$. The area of the triangle A is the total materials tied up in work-in-process inventories, and B is the corresponding value-added.

Equation (5) provides a rough first approximation for the production period. How good an approximation it is depends on the relevance of many of the assumptions used. Some considerations suggest that x computed from (5) will overstate the actual production period and others that x will be too low. For example:

1). If most materials enter at the beginning of the productive process, instead of evenly throughout, then equation (5) overstates the production period. The extreme case of all materials being added at the start of the process is depicted in Figure 1b. Remember that W_t (the area of A and B combined) and P_{t+1} are given and that x is to be determined. Abramovitz, p. 174, has also pointed out this possibility.

2). If a portion of value-added is a markup that enters into the gross value of production when the items become finished goods but not into the estimates of work-in-process inventory, then equation (5) underestimates the production period. This is depicted in Figure 1c. Again x must be adjusted so that area of A and B combined is the same as in Figure 1a.

3). If production is growing along a linear trend, then $2W_t/P_{t+1}$ will underestimate x slightly if x is less than 1.5. The reverse holds for declining production.[2]

4). There may be "stages" of processing between which work-in-process stocks are temporarily stored. In such cases, equation (5) will underestimate the production period if most of the processing occurs at the later stages.[3] Other things equal, of course, the longer the wait between stages, the longer will be the overall production period.

Formula (5), to repeat, is only a rough approximation, subject to vagaries of accounting systems and peculiarities of different productive arrangements. However, it does give some indication of the order of magnitude of the production period in varying industries and the numbers presented below do have an intuitive plausibility. Perhaps more experienced observers of manufacturing operations can verify or cast doubt on this initial impression.[4]

Data on work-in-process inventories and levels of production are available from the Bureau of the Census. The Annual Survey of Manufactures provides year-end stocks of inventory by stage of fabrication as well as cost of materials consumed and value-added in manufacture for two, three, and four-digit industries.

Monthly data on inventories are also available for two-digit industries. Only a few of these published series show inven-

[2] This result follows from assuming that production changes at a linear rate r per period:

$$q(t-u+x) = q(t)[1 + r(x-u)]$$

In that case from (4):

$$P_{t+1} = q(t)(1 + r/2)$$

and from (3), with the assumption that $m(y)+v(y) = 1/x$:

$$W_t = q(t)(3x + rx^2)/6$$

Then

$$\frac{2W_t}{P_{t+1}} = \frac{x(3 + rx)}{3 + 3r/2}$$

which for positive r is greater than x if x is greater than 3/2.

[3] To see this for the two-stage case illustrated in Figure 1d, let b = time between stages, s = time to complete the first stage and x = total production period. Actual processing time is $x-b$. Either by integration or by direct computations of areas in the figure, we have

$$W_t = [(x-b)/2 + sb/(x-b)]P_{t-1}$$

Rearranging terms

$$\frac{2W_t}{P_{t+1}} = x + b\left(\frac{2s}{x-b} - 1\right)$$

Therefore, $2W_t/P_{t+1}$ is less than x if the first stage is completed in less than half the actual processing time, i.e., if $s<(x-b)/2$.

[4] I have used accounting data from a bag manufacturing plant and found that the resulting estimate was very close to the production manager's guess as to the length of the production process in the plant.

tories by stage of fabrication, but monthly Census data have been prepared for all two-digit industries from 1953 to 1966 by David Belsley, who has generously made these data available. For a description of the data see Belsley's book.

II. Production Periods and the Stock-Order Dichotomy

In a study of the German leather and shoe industries T. Thonstad and D. B. Jochems write, "... we may reasonably assume that the tanners are more engaged in production for stock building than are the shoe manufacturers, since their period of production is so much longer" (p. 143). Before considering their implicit hypothesis about stocks and the production period we shall examine whether or not our formula using *U.S.* data reveals a longer production period for tanners than for shoe manufacturers.

Industry 3111, Leather Tanning and Finishing, and industry 3141, Footwear, Except House Slippers and Rubber Footwear, should correspond reasonably well with the two industries Thonstad and Jochems studied. We considered data for 1964–1966 from the Survey of Manufacturers. For each year average monthly production was set equal to the sum of cost of materials consumed and value-added, adjusted, divided by 12. We added the end-of-year work-in-process inventories for the current and previous years and then divided this figure by average monthly production to get an estimate of the production period. The results in months are as follows:

	Industry 3111 Tanning	Industry 3141 Shoes
1964	1.84	.54
1965	1.76	.53
1966	1.68	.54

The figures suggest that tanners have a production period about three times as long as the one for shoe manufactuers. We cannot tell to what extent there is a seasonality in year-end inventory that might distort the results. For industry 31 as a whole there is no evidence of a notable year-end change in W_t, but this does not tell us much about industries 3111 and 3141. The estimates, however, are definitely consistent with Thonstad and Jochem's observation that tanners have a much longer production period than shoe manufacturers.[5]

Is it plausible to expect that the length of the production period would be related to the extent to which an industry produces for stock rather than to order? In the quoted passage, Thonstad and Jochems seem to imply that the longer the production period, the more there is production to stock. The decision about which way to organize production is likely to depend on the relative costs of holding inventories versus costs of backlogging orders. (For example, see Holt et al. or Carlson and T. B. O'Keefe.) In the case of a reasonably standardized product like leather, the potential cost of having unsaleable stock is not so great as in an industry such as shoe manufacturing where styles may become obsolete. Thus, any tendency for tanners to produce more to stock than shoe manufacturers may be attributable to such relative costs rather than to the length of the production period.

In a world of certainty, the production period should have no relationship to the stock-versus-order distinction. Except for considerations of efficient batch size or production-smoothing, there should be neither stocks of finished goods nor unfilled orders. Production can be begun

[5] Ruth Mack, for illustrative purposes, suggests a two-month period for shoe leather and a one-month period for upper leather when constructing a "minimum service" stock for tanners, p. 192. Our estimates for the tanning industry are clearly of the same order of magnitude.

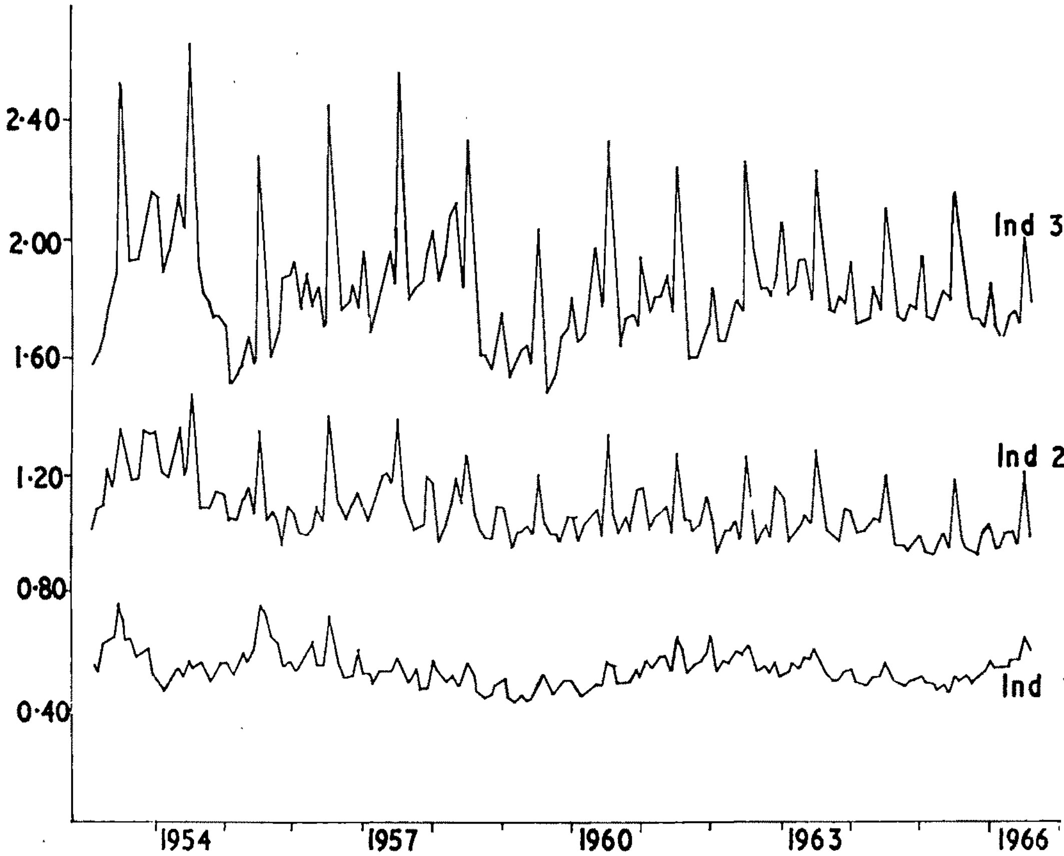


FIGURE 2. MONTHLY CALCULATIONS OF $2W_t/P_{t+1}$ FOR INDUSTRIES 27, 22, and 36.

sufficiently in advance so that finished goods will be ready precisely when demanded. In a world of uncertainty, as noted before, whether production is in anticipation of demand or in response to demand depends to a large extent on relative expected costs. If items that have long production periods are generally complicated and built to varying specifications, then they will also tend to be produced to order. Thus there may be an empirical relationship between the production period and the relative size of unfilled orders that is not intrinsically a result of the length of the production time.[6] However, if orders are backlogged, then the longer the production period, the longer we would expect the delivery period to be.

[6] For a recent discussion of reasons for having unfilled orders, see Gerald Childs who also summarizes some of Abramovitz' ideas on the subject.

The production period for two-digit industries can be estimated from two different sets of data. With Belsley's data, not seasonally adjusted, the ratio $2W_t/P_{t+1}$ in each industry has been computed for every month from February 1953 to August 1966. For illustrative purposes, the results for three industries are graphed in Figure 2. The arithmetic mean of the production period in each industry for each full year 1954 to 1965 has also been computed. The figures for 1965, with the sample standard deviation of the mean in parenthesis, are shown in column 2, Table 1. Alternatively, data from the Survey of

TABLE 1—1965 INDUSTRY PRODUCTION PERIODS AND DELIVERY PERIODS[c]
(All entries are measured in months)

	Average Production Period x (1)	(2)	Ratio of Finished Stocks to Shipments (3)	Average Delivery Period d (4)	$\frac{d-x}{2}$ (5)
Industries without unfilled orders					
20 Food	.14[a]	.14[b] (.001)	.55		
21 Tobacco	.15	.25 (.007)	.30		
23 Apparel	.58	.60 (.005)	.87		
28 Chemicals	.39	.41 (.002)	.69		
29 Petroleum	.45	.47 (.001)	.64		
30 Rubber, Plastics	.47	.38 (.003)	.73		
Industries with unfilled orders					
22 Textiles	.97	.97 (.006)	.82	1.05	.05
24 Lumber	.71	.51 (.004)	.69	.61	.05
25 Furniture	.78	.89 (.006)	.48	.95	.03
26 Paper	.28	.29 (.001)	.42	.47	.09
27 Printing	.46	.48 (.001)	.35	.22	—
31 Leather	.73	.66 (.003)	.63	.81	.08
32 Stone	.38	.44 (.002)	.90	1.06	.31
33 Primary Metals	1.07	1.16 (.007)	.52	1.79	.32
34 Fab Metals	1.32	1.47 (.006)	.61	2.49	.51
35 Machinery	2.31	2.36 (.013)	.78	2.95	.30
36 Elect. Mach.	1.85	1.81 (.011)	.57	3.14	.67
37 Trans. Equipment	1.70	1.76 (.018)	.19	4.09	1.18
38 Instruments	1.83	2.44 (.011)	63	2.18	—

[a] Computed from Annual Survey of Manufactures, 1966, General Statistics for Industry Groups and Industries, and Value of Manufacturers' Inventories.
[b] Computed from M3 Monthly Survey data by the Bureau of Census, data prepared for David Belsley. Columns (3), (4), and (5) from same source.
[c] See text for method of calculating delivery periods.

Manufactures yield the estimates shown in column 1. The two sets of estimates are similar.

The first thing to note is that there is no strict dichotomy separating the average production period of industries with unfilled orders from those without. There is evident overlapping.

It is also apparent that on balance, industries without unfilled orders do tend to have lower production periods. We can check this observation by performing a median test as described by S. Siegel, pp. 111–15. An industry is classified according to whether or not it has unfilled orders and whether its estimated production period is above or below the median for all industries. The following array results with the figures from column 2 of Table 1:

	Industries without unfilled orders	Industries with unfilled orders	Totals
Production period equal to or above median	1	9	10
Production period below median	5	4	9
Totals	6	13	19

For 1965, industry 23 has the median production period. This has been included in the upper part of the array in order to get the greatest probability of finding this or a more extreme distribution on the assumption that both categories have the same median. This probability is slightly less than .05. Since in every other year, as well as with column 1 figures, all six indus-

tries without unfilled orders have production periods below the combined median, we reject the hypothesis of a common median and tentatively claim that industries without unfilled orders do tend to have shorter production periods.

III. Delivery Periods and Stock-Versus-Order Production

Consider next the overall reaction time within manufacturing industries and how the production period influences its length. Assume initially that custom, technology, and least-cost considerations have determined whether an item is produced to order or to stock. The reaction time can then be analyzed for each situation separately.

Production to Order

When production is to order, the initiating signal for a manufacturing firm is a new order. The length of time from receipt of the order until shipment to fill the order is called the *delivery period* or lead time.

It is convenient for us to divide the delivery period d into three phases: waiting time, incubation time, and accumulation time. These are depicted in Figure 3a. The waiting time w, from t_0 to t_1, is the interval from receipt of the order until the beginning of its production. The incubation time from t_1 to t_2 is the production period x. Finally, there may be an accumulation time c from t_2 to t_3 in which an order is completed. Thus

$$d = w + x + c \tag{6}$$

The delivery period undoubtedly depends on capacity constraints. If capacity is fully utilized when an order arrives, then the initiation of production must wait. The larger the backlog of orders relative to capacity, the longer will be the waiting time w. The accumulation time c may also be affected by capacity constraints. With excess capacity, several items for an order

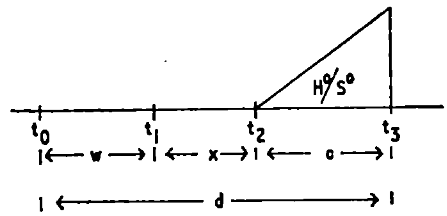


FIGURE 3a. PRODUCTION TO ORDER

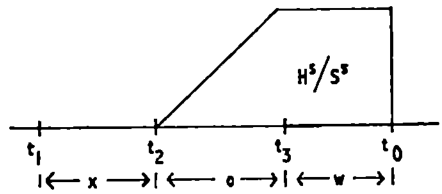


FIGURE 3b. PRODUCTION TO STOCK

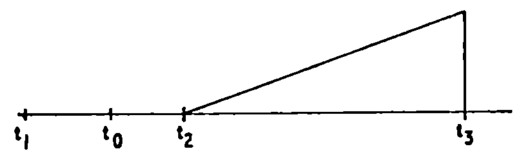


FIGURE 3c. MIXED CASE

FIGURE 3. GRAPHICAL ILLUSTRATION OF AVERAGE FINISHED INVENTORY PER UNIT OF SHIPMENTS.

t_0 = TIME ORDER IS RECEIVED
t_1 = TIME PRODUCTION IS STARTED
t_2 = TIME FINISHED GOODS START TO ACCUMULATE
t_3 = TIME PRODUCTION IS COMPLETED ON THE LOT

may be produced simultaneously while with tight capacity the items may be worked on sequentially and the accumulation time strung out.

The size of each order can also influence the accumulation time. If each order is for a single item, there will be no accumulation time. With many items in an order, the manufacturer may accumulate the entire order before shipment and hence hold inventories against backlogged orders. Abramovitz, p. 404, has estimated that such stocks comprise 15–25 percent of all finished goods in manufacturing industries.

Suppose finished goods flow out of the manufacturing process at a uniform rate and that as soon as one batch of items is shipped another begins to accumulate. In such a case, the average stock of inven-

tories per unit shipped is:

$$(7) \quad H^0/S^0 = c/2 = (d - x - w)/2$$

where H^0 denotes finished inventories held for unfilled orders and S^0 shipments to fill unfilled orders. This formula is depicted in Figure 3a where the height of the triangle is one unit.

Production to Stock

When production is to stock, the initiating signal occurs when production begins. There is again an incubation period, before finished goods are completed, and an accumulation period. The waiting period, however, comes at the end while the items await an order.[7] If shipment occurs immediately upon receipt of an order, the average stock of finished inventories per unit shipped is

$$(8) \quad H^s/S^s = w + c/2 = (d - x + w)/2$$

where S^s denotes shipments from stock and H^s the corresponding stock. Formula (8) is depicted in Figure 3b. The term d denotes the delay from initiation of production until shipment and is analogous to the delivery period but its duration is not so readily estimated from available data.

The Delivery Period

Victor Zarnowitz, p. 16, has suggested that the ratio of unfilled orders to shipments, in the absence of any better measures, may be used as a crude indicator of an industry's average delivery period. With available data there is no way of knowing what proportion of shipments go to fill backlogged orders and what proportion is out of stock. Zarnowitz's indicator, however, is in a sense independent of these proportions. To see this let

S = total shipments during a period
U = unfilled orders at the beginning of the period
r = proportion of S that fill backlogged orders ($r = S^0/S$)

Assuming r and S are reasonably steady, then U/rS is the average delivery period for items produced to order. The average delivery period on the rest of the shipments is zero, ignoring response time in filling orders from stock. If these two delivery periods are averaged by assigning a weight r to U/rS and weight of $1-r$ to 0, the average is evidently U/S. With monthly data, this indicator, if greater than one, could be refined a bit more by using shipments in subsequent months, but for our purposes this refinement is not essential. Thus, the figures labelled Average Delivery Period in Table 1 are obtained by dividing the sum of unfilled orders in an industry from December 1964 to November 1965 by the total shipments during 1965.

With our rough estimates of the production period and the delivery period, we can investigate the possibility in industries reporting unfilled orders that *all* production is to order. The figures for d and x in Table 1 immediately rule out such a possibility for industries 27 and 38, for if all production is to order, then d must be at least as large as x. To obtain the average delivery period just for items produced to order, the relevant ratio is U/S^0. In industries 27 and 38, the proportion of production for stock must be at least 55 and 12 percent, respectively, if the delivery period on items produced to order is to be no shorter than the average production period.

Equation (7) can be used to examine whether finished stocks in industries with unfilled orders could be the result of pro-

[7] As a referee has suggested, there are undoubtedly prior considerations, such as desired inventory-to-sales ratios, that determine the planned rate of production. Variations in the average waiting time would then depend in part on the ability of decision makers to recognize the relationship between the planned and the actual situation.

duction to order or whether they result primarily from production to stock. Consider Figure 3a. We have estimates of d and x but not of w and c. Given d and x, the smaller is w, the larger is c, and hence, the larger is H^0/S^0. Subject to a qualification below, we can put an upper limit on H^0 by assuming that all production is to order and that w is zero. In such a case, from equation (7), H^0/S^0 becomes $(d-x)/2$. These figures are shown in column (5) of Table 1.

If $(d-x)/2$ is viewed as an upper limit on the ratio of finished stocks to shipments when production is entirely to order, it becomes apparent when comparing columns (3) and (5) of Table 1 that in industries 22, 24, 25, 26, and 31 there must be a substantial amount of production to stock. Similar interpretations hold for industries 32–35 although in these industries a larger portion of the finished stocks could be the result of accumulation of goods produced to order. Only in industry 36 and most notably in 37 does $(d-x)/2$ exceed the actual ratio of finished inventories to shipments. Without more direct information, it seems believable from these figures that most production in industry 37 is to order. If so, with a ratio of finished stocks to shipments of .19, the implied value of waiting time w is 1.97 months and of accumulation time c is .36 months.

Unfortunately, this is about as far as this approach can go in separating shipments into their stock-versus-order components. In each industry, except possibly 37, there is undoubtedly production to stock. However, even if we had a guess as to the size of H^s, we cannot use equation (8) to determine S^s without more information about the values of x, c, and w for items produced to stock. Such a wide range of values is plausible that almost any proportion $r=S^0/S$ from 0 to 1 could be obtained.[8]

Belsley for computational convenience assumes that $H^0\equiv 0$, but stresses in a footnote that this is not to be taken as a behavioral hypothesis, p. 62. We question the assumption for industry 37 and it may not be a very good approximation for industries 32–36. In the remaining industries it appears to be readily justifiable. Column (5) of Table 1 was intended to give the maximum H^0/S on the assumption that all production was to order. H^0 may thus be considerably lower than the numbers indicate. There is, however, another possibility shown in Figure 3c. Production may begin prior to receipt of an order and then completed to specification after the order is received. In this case the delivery period would be from t_0 to t_3 but the accumulation period could be much larger than $d-x$ and thus H^0/S could be larger than $(d-x)/2$. Without more information little else of substance can be said on the relevance of Belsley's assumption that $H^0\equiv 0$.

Since the production period is one component of the delivery period, we would expect to find a positive association between the two. To check this hypothesis, the industries with unfilled orders were ranked by size of production period and by size of delivery period. The Kendall rank correlation coefficient is .67, which is highly unlikely under the null hypothesis of no association. The t-statistic is equal to 3.[9]

IV. Changes in the Production Period

Holt and Modigliani discuss three types of productive processes. They are:

[8] Belsley, using a different approach, was also unable to determine how much of the industry aggregates fell into each of these two categories even though he did finesse the problem ingeniously. He later commented: "It is not too surprising, although disappointing, that we are unable to unravel from these few short-run parameter estimates enough information to put back together what was never there in the first place" (p. 146).

[9] The computational procedure for obtaining these statistics is presented by Federick C. Mills, pp. 312–17.

Type A: Production time decreases as rate of production is increased
Type B: Production time is not affected by rate of production
Type C: Production time increases as rate of production increases

They suggest that the "... type C form of operation is the most general and perhaps the most common one" (p. 40). With our estimates of the production period, we can examine this conjecture, although it should be borne in mind that by construction, the estimate of the production period is likely to fall as production rises since production is in the denominator. If the reverse holds, we would interpret such a result as strong confirmation of Holt and Modigliani's conjecture.

A cursory inspection of our production period over seasonal fluctuations gives a strong impression that most industries behave in accordance with type A, but this may be the result of the manner in which the production period is computed.

An alternative approach is to examine the relationship between $\Delta P/P$ and $\Delta W/W$ where it is understood that the percentage change in W is taken one period earlier than the corresponding percentage change in P. The following equation then holds approximately

$$\frac{\Delta x}{x} = \frac{\Delta W}{W} - \frac{\Delta P}{P} \tag{9}$$

where x is the production period as estimated by equation (5).

Suppose there is a linear relationship between the percentage changes in P and W:

$$\frac{\Delta P}{P} = a_1 + b_1 \frac{\Delta W}{W} \tag{10}$$

Substituting in (10) for $\Delta W/W$ from (9) and solving for $\Delta x/x$ yields:

$$\frac{\Delta x}{x} = -\frac{a_1}{b_1} + \frac{(1 - b_1)}{b_1} \frac{\Delta P}{P}$$

If in estimating equation (10), we find that b_1 is greater than 1, then we have evidence that x and P move inversely and that production is of type A.

If b_1 is less than 1 we cannot immediately conclude that production is of type C. It may be that the relationship is not sufficiently close to conclude anything. However, we can reverse the regression and estimate

$$\frac{\Delta W}{W} = a_2 + b_2 \frac{\Delta P}{P} \tag{11}$$

If b_2 is greater than 1, then we could say that production tends to be of type C.[10]

The results of estimating equation (10) for each manufacturing industry over the period 1953–66 are shown in Table 2. There are three industries (22, 35, and 37) which given an indication that production is of type A in that the coefficient b_1 is greater than 1.

With the other regressions none of the b_2 coefficients was greater than 1. These are not reported but can be calculated from the entries in Table 2 by using the relationship $b_2 = b_1 s_2^2/s_1^2$ where s_1 and s_2 are the standard deviations of $\Delta P/P$ and $\Delta W/W$, respectively. With the exception of industries 20 and 29, the s_1 and s_2 figures indicate that reported production tends to vary relatively more than reported work-in-process inventories. Except for the two cases noted, b_2 is smaller than b_1 and in those two cases b_2 is not appreciably greater than b_1. Although there are only three industries for which we are prepared to rule out the possibility that production is primarily of type C, there is certainly no real support for Holt and Modigliani's conjecture that type C is the most common form of operation.[11] As noted above both

[10] Because of the aggregative nature of the data, we have not formulated this as a precise problem in hypothesis testing but rather as a tentative exploration.

[11] There may be a semantic difficulty here. With stages of work in process, our measure of the production

TABLE 2—INDUSTRY ESTIMATES OF THE REGRESSION EQUATION $(\Delta P/P) = a_1 + b_1(\Delta W/W)$

Industry	a_1	b_1	t_b[c]	s_1[a]	s_2[b]
20	.0048	.027	.41	.046	.055
21	.0410	.452	.92	.382	.061
22	.0053	1.275	3.74	.111	.025
23	.0100	.364	2.32	.141	.070
24	.0074	.422	3.57	.093	.060
25	.0057	.692	3.01	.119	.040
26	.0091	−.291	−1.55	.069	.029
27	.0051	.386	4.30	.075	.062
28	.0060	.139	.72	.061	.025
29	.0034	.092	1.25	.032	.035
30	.0069	.564	3.43	.080	.037
31	.0065	.363	2.20	.094	.044
32	.0036	.475	3.92	.075	.047
33	.0103	.332	.87	.143	.029
34	.0044	.317	1.52	.080	.030
35	−.0008	1.700	4.48	.085	.017
36	.0084	.405	.74	.113	.016
37	.0005	1.473	4.37	.102	.023
38	.0040	.921	2.67	.092	.021

[a] Standard deviation of $\Delta P/P$
[b] Standard deviation of $\Delta W/W$
[c] t-statistic for the estimate of b_1

the waiting time and accumulation time are likely to increase as production levels get closer to capacity; but if anything the evidence suggests that individual items get completed more rapidly.

In 18 of the 19 industries, there are positive b coefficients. In 11 cases, these are significantly different from zero at a .95 confidence level. The column headed t_b gives the t-statistics for the b coefficients. Thus, production does tend to vary in the same direction as work-in-process stocks at the end of the previous month. This suggests taking a closer look at the relationships between W and P.

period would become shorter if there were less time between stages. If by production time, Holt and Modigliani are referring to something like actual processing time, $x-b$ in Figure 1d, then the aggregate data cannot be used to determine what happens. If they mean x, the production period as we have developed the concept, then there is definitely no evidence of type C operations in the two-digit industry data.

V. An Attempt to Extract Additional Information From the Data

Formula (3) in Section 1 indicates that work-in-process stocks are a distributed-lead function of future production. Thus, a regression of work-in-process at the end of month t (W_t) on production in future months ($P_{t+1}, P_{t+2}, \ldots$) might be able to pick out roughly the extent to which work-in-process is written up as a function of the time until completion of production.

We tried regressions using nonseasonally adjusted data with a variety of lead relationships.[12] The general conclusion, with these data at the level of two-digit industries, is that variations in future produc-

[12] These were run by Robert Meyer, who has developed a working computer program to estimate distributed leads and lags using frequency domain relationships. I am indebted to him for overseeing the actual computations and for helping to interpret the results. The technique requires putting no a priori restrictions on the relationships among the coefficients. For additional information, see the cited papers by Thomas F. Cargill and Meyer.

TABLE 3—DISTRIBUTED LEAD ESTIMATES OF b AND c IN

$$W_t = a + bP_{t+1} + cP_{t+2}$$

Using Monthly Two-Digit-Industry Data 1953–66

Industry	b^{*a}	c^{*a}	b	Ratio[b]	c	Ratio[c]	R^2
20	.07		.004	.87	−.008	−1.86	.024
21	.10		.003	1.00	.001	.51	.006
22	.52		.030	3.84	.006	.71	.069
23	.28		.028	3.55	.027	3.46	.095
24	.36		.074	4.28	.024	1.37	.091
25	.42		.040	4.05	.010	1.02	.071
26	.14		−.013	−3.62	−.006	−1.75	.058
27	.27		.064	4.64	.015	1.11	.110
28	.21		.006	1.09	.011	1.78	.023
29	.26		.014	.78	−.012	−.64	.009
30	.24		.025	4.16	.005	.74	.083
31	.39		.054	5.06	.053	4.90	.130
32	.20		.027	4.45	.042	6.94	.214
33	.55	.01	.019	3.13	.006	.90	.037
34	.64	.05	.042	2.42	.020	1.15	.024
35	.78	.35	.092	7.51	.002	.19	.195
36	.72	.19	−.004	−.63	−.042	−6.14	.129
37	.76	.26	.052	3.32	.013	.82	.057
38	.80	.39	.053	2.97	−.014	−.79	.069

[a] Estimates of b^* and c^* are predicted coefficients on basis of assumptions used in constructing average production period. See equation (5) in text.
[b] Ratio of coefficient b to its standard error.
[c] Ratio of coefficient c to its standard error.

tion do not explain much of the variations of work-in-process. Table 3 presents the results when it is assumed that current work-in-process is determined by production in the following two months.

The columns labelled b^* and c^* were prepared as if the assumptions used in constructing the production-period formula $x = 2W/P$ all held. That is, if production takes precisely a length of time x and work-in-process is written up linearly over the entire period, then it is possible to predict what values the coefficients will take. In industries in which the average production period is less than a month, the predicted value of b was taken to be average work-in-process $\overline{W}$ over average production $\overline{P}$, the averages taken over 163 monthly observations. When according to the formula the average production period is more than a month, the ratio $\overline{W}/\overline{P}$ was allocated to future months by the proportion of the work-in-process triangle, such as in Figure 1a, that falls in each of the future months. This accounts for the entries in the column c^* for industries 33–38.

The regression coefficients b reported in Table 3 are all much lower than the coefficients that would arise if all the assumptions behind the production-period formula held. All but two of the coefficients are positive and thirteen of them are significantly so (i.e., the ratio of the coefficient to its standard error is greater than 2). The results for c are more erratic, as were the coefficients for other values of P in the various distributed-lead regressions that were run.

While there is some association of the sort predicted, we are left with the impression that the regressions have not picked up the full relationship that probably exists between work-in-process and future production. At this point we can only offer

a number of conjectures as to why.

1. The accounting records may not be kept with sufficient precision.

2. There are a number of reasons, cited by Thomas Stanback, Jr., pp. 95–99, why production data will neither be coincident with, nor lag behind work-in-process by a given amount with any regularity. One of the more important possibilities is that there may be considerable variation in levels of goods in process between stages of fabrication. As orders increase, a number of semifinished items could be finished rapidly so that it is possible to have, in the short run, a negative relation between goods in process and production of finished goods.

3. There are undoubtedly serious problems of aggregation. For one thing, firms with very different production times are included in the same industry. More importantly one manufacturing plant's variations in production and work-in-process might offset another's variations in ways that in the aggregate could mask or distort the link between the series that might exist for each plant separately.

The empirical importance of these aggregation effects could be determined only with reliable data from a variety of manufacturing plants, perhaps from those that comprise the sample from which the aggregates themselves have been constructed. With access to such data, it might be possible to ascertain more about the distribution of production periods and the appropriate assumptions about the relationship between work-in-process stocks and production in different industries. Lacking access, we must arbitrarily assume the relationship as we did in making the rough estimates of the average production periods in each industry.

We also ran regressions between production and new orders with the intent of determining whether or not there were any consistent lead or lag relationships. The only consistent pattern, and it was very strong, was of coincident movements between the two series. There are a number of possible explanations. First, the production series is linked by construction to the new-orders series. In fact, production equals new orders minus changes in unfilled orders plus changes in finished goods inventory. Second, the data were not seasonally adjusted so that seasonal demands may be reflected in production plans geared to expected new orders before the orders actually arrive. Third, there may be effects of the sort mentioned above in which changes in between-stage goods in process may be utilized when necessary to have production conform fairly closely to new orders.

VI. Summary and Conclusions

A very simple formula has been used to estimate the production period in manufacturing industries. With these estimates, we may conclude:

1. The average production period for all two-digit *U.S.* manufacturing industries is less than a quarter of a year and for two-thirds of them it is less than one month. While adjustments in rates of production are not entirely dependent on the length of the production period, the order of magnitude of these figures provides support for incorporating, in models of inventory behavior, parameters to represent within-quarter adjustment of production rates in response to unanticipated levels of sales. See, for example, Albert Hirsch and Michael Lovell, p. 204. At lower levels of aggregation there is, of course, much greater dispersion between estimates of industry production periods.

2. Industries that report no unfilled orders and hence are presumed to produce solely to stock tend to have lower production periods than industries that have some production to order.

3. Among industries that produce to

order there is, as would be expected, a strong positive association between the length of the production period and the length of the delivery period.

4. Estimates on the production period and the delivery period can provide only limited and highly imprecise inferences about what proportion production to stock is of total production in an industry. They can be used, however, to cast doubt on Belsley's operational assumption that all industries carry finished goods only of items produced to stock. In one industry the figures suggest that there may be relatively little production to stock by the establishments reporting to the Census survey.

5. There is no evidence in support of Holt and Modigliani's conjecture that production time increases as the rate of production increases. If anything, the evidence suggests just the opposite.

These tentative results are of sufficient interest, it would seem, to justify further exploration with estimates of the production period by means of accounting data, particularly if adequate series could be utilized at a much more disaggregated level than that of two-digit industries.

REFERENCES

M. Abramovitz, *Inventories and Business Cycles,* New York 1950.

D. A. Belsley, *Industry Production Behavior: The Order-Stock Distinction,* Amsterdam 1969.

E. v. Bohm-Bawerk, *The Positive Theory of Capital,* New York 1923. (Translated from the 1888 German edition by William Smart in 1891.)

T. F. Cargill and R. A. Meyer, "A Spectral Approach to Estimating the Distributed Lag Relationship between Long and Short Term Interest Rates," *Int. Econ. Rev.,* June 1972, *13,* 223–38.

——— and ———, "A Simulation Study of the Hannan Procedures for Estimating a Distributed Lag Process," *Proc. Amer. Statist. Ass. 1971,* vol. 12, Washington.

J. A. Carlson and T. B. O'Keefe, "Buffer Stocks and Reaction Coefficients: An Experiment with Decision Making under Risk," *Rev. Econ. Stud.,* Oct. 1969, *36,* 467–84.

G. L. Childs, *Unfilled Orders and Inventories,* Amsterdam 1967.

A. A. Hirsch and M. C. Lovell, *Sales Anticipations and Inventory Behavior,* New York 1969.

C. C. Holt and F. Modigliani, "Firm Cost Structures and Dynamic Responses of Inventories, Production, Work Force, and Orders to Sales Flunctuations," in Joint Economic Committee *Inventory Fluctuations and Economic Stabilization,* Part II, Washington 1961 3–55.

———, ———, **J. F. Muth, and H. A. Simon,** *Planning Production, Inventories and Work Force,* Englewood Cliffs 1960.

R. P. Mack, *Consumption and Business Fluctuations: A Case Study of the Shoe, Leather, Hide Sequence,* New York 1956.

F. C. Mills, *Statistical Methods,* 3d ed, New York 1955.

S. Siegel, *Nonparametric Statistics for the Behavioral Sciences,* New York 1956.

T. M. Stanback, Jr., *Postwar Cycles in Manufacturers' Inventories,* New York 1962.

T. Thonstad and D. B. Jochems, "The Influence of Entrepreneurial Expectations and Appraisals on Production Planning: An Econometric Study of the German Leather and Shoe Industries," *Int. Econ. Rev.,* May 1961, *2,* 135–153.

V. Zarnowitz, "Unfilled Orders, Price Changes, and Business Fluctuations," *Rev. Econ. Statist.,* Nov. 1962, *44;* reprinted as Nat. Bur. Econ. Res. occas. pap. 84, New York 1962.

The Not-For-Profit Hospital as a Physicians' Cooperative

By Mark Pauly and Michael Redisch*

The private, nonprofit hospital has usually been regarded by economists as an organizational anomaly. In particular, it has been alleged that the predominance of the not-for-profit structure within the American hospital system is associated with a weakening of the usual market constraints of competition and profit orientation. As a result, this critical element in any analysis of the medical care system in the United States has usually been modeled with a mixture of anecdote and *ad hoc* assumption. It is typically assumed that "all objectives of nonprofit organizations can be described in terms of some type(s) of output (broadly defined) or capital stock."[1] William Baumol and Howard Bowen describe these goals as "bottomless receptables into which limitless funds can be poured" (p. 497).

Model variation occurs as investigators place combinations of key variables in either the objective function or the constraint set of the hospital. Joseph Newhouse (1970) and Martin Feldstein (1971) studied the implications of the maximization of quantity-quality subject to a budget constraint. Millard Long's model is one of quantity maximization subject to both a budget and a quality constraint. Paul Ginsburg assumed maximization of weighted output subject to a budget and an availability of capital constraint. Maw Lin Lee included types of physical capital in the hospital objective function. And Melvin Reder talked of hospitals trying to maximize "the weighted number of patients treated (per time period), the 'weights' being the professional prestige to the doctors attending them" (p. 480).

This last model is the only one to (even) hint at a nonpassive role for the physician in a model of hospital behavior.[2] In this paper we propose an alternative model in which the physician emerges as a traditional income maximizing economic agent who is "discovered" in a decision-making role within this not-for-profit enterprise. Our model is similar to the model of the firm customarily used by economists, in that it is based on the assumption of net income maximization. Only a somewhat unusual definition of net income is needed to enable us to apply in our short-run analysis many of the conclusions of the orthodox model of the firm. In the longer run, however, our model, while still based on net-income maximization, yields different predictions about the institution's response to changes in demand and supply parameters. Furthermore, it may be possible to generalize parts of our model to other private, nonprofit service firms such as universities and symphony or-

* Northwestern University and the Office of Health Policy Analysis and Research, Department of Health, Education, and Welfare, respectively. Research support was provided by the Health Services Research Center of Northwestern University and the American Hospital Association and the Health Economics Research Center of the University of Iowa. Portions of this paper were presented at the Health Economics Session of Allied Social Sciences Meetings, December 1969. Helpful comments were received from R. Tollison, C. M. Lindsay, and James Buchanan. Views expressed are those of the authors.

[1] See Paul Ginsburg, p. 42.

[2] Paul Feldstein and Carl Stevens have discussed the role of the physician in the hospital, but have not provided an explicit model.

chestras, which produce and sell services to individual consumers, many of whom are not poor.

Specifically, we assume that the group of attending physicians on the hospital's staff enjoys *de facto* control of the hospital at any point in time. Given this assumption, we develop a model in which the hospital operates in such a way as to maximize the net income per member of the physician staff. Results are obtained which are similar to those derived from models of producers' cooperatives in Yugoslavia and collective farms in the *USSR*. The physician plays a role analogous to that of the Yugoslav worker and the Russian peasant. Our results are also similar to those obtained from the "theory of clubs," developed by James Buchanan and others.

I. The Model

We simplify the problem initially by assuming that patients pay the full market price for care, and that the decision-making group in the hospital is able to impose its collective will on individual members. The implications of weakening these assumptions to allow for customary forms of health insurance and for imperfect cooperation among controlling individuals will be discussed in later sections.

The product produced in the hospital is hospitalization services. We shall assume that this output can be represented by a single variable Q.[3] To produce this output, physical capital (K), nonphysician labor (L), and physician or medical staff labor (M) are used. The production process can be summarized by the production function

$$Q = F(K, L, M) \tag{1}$$

In European countries in which physicians who treat patients in the hospital are employed by and paid by the hospital, this three-input production function is the obvious one.[4] But in the United States, the hospital patient is subject to two separate billings. The hospital charges him only for the use of capital and nonphysician labor. The physician presents a separate bill for the use of his "personal" physician's services. This dual billing system has led to a conceptually false dichotomy in much of the health economics literature. The physician and hospital are often viewed as independent economic entities selling services in functionally segmented health markets. This view appears to provide the rationale for the hospital-administrator-oriented, output-maximization theories of hospital behavior discussed earlier.

We propose an alternative view. It seems obvious that the patient's demand is primarily for the service produced by the physician and hospital acting in combination, not for the separate components, even though there probably is a separate demand for some attributes, such as amenities, or additional patient days for recuperating, that the hospital alone produces.

The critical assumption of our model is that the physician staff members enjoy *de facto* control of hospital operations and see to it that hospitalization services are produced in such a way as to maximize their net incomes. The appearance of physician control is not hard to establish. The staff physicians have direct control over the number and types of patients admitted and over the types of services

[3] Derivatives of two quite different surrogates for hospital output have been used most often by economists doing empirical research. One is based on the number of inpatient days and outpatient visits at the hospital while the other is concerned with the number of cases treated in the hospital. The "case treated" corresponds most closely with the measure of output implied by our model.

[4] This was the form used by M. Feldstein (1967) in in his study of hospitals in the United Kingdom.

they receive; they control output. The staff physicians can determine, within rather broad limits, what use of the hospital will be made in treating a patient; they control many of the production decisions. They have indirect control over many other aspects of the hospital's operation, such as capital investment and the level of nursing care, in the sense that no administrator can afford to incur the displeasure of the medical staff, interfere with medical staff prerogatives, or make decisions which will deter large numbers of physicians from remaining on the hospital's staff or using that hospital for their patients. The trustees, who have nominal control over the hospital's operation, usually look to the medical staff when making decisions on operations or capital investment.

We first assume that the physicians on the staff of a hospital at any point in time act in such a way as to maximize the sum of the money incomes of all staff members. Such an assumption implies a process of group decision making resulting in a kind of perfect cooperation not likely to be observed in practice. It also ignores nonmonetary components of a physician's income, such as leisure time and prestige, which are likely to be of some importance. Nevertheless, this model is useful as a bench mark from which to consider the effects of alternative assumptions.

We postulate an economic short-run period as one in which the number of physicians on the hospital staff, M, remains constant. Each physician is presumed to supply a constant, homogeneous amount of medical input.

The patient's demand is primarily for the service produced by the physician and hospital acting in combination, not for the separate components. This can be formalized by postulating a demand function for "hospitalization services" faced by the physician staff that takes the form

$$(2) \qquad Q = Q(P_T), \qquad \frac{\partial Q}{\partial P_T} < 0$$

where P_T is the combined price paid by the patient for the physician and hospital components.[5,6]

We also assume that the hospital component of P_T is set so as to allow the hospital to just break even, with no gain or loss.[7] That is, we assume that the hospital price P_h is set to produce the equality:

$$(3) \qquad P_hQ = wL + cK$$

where P_h is the unit price the hospital charges for use of nonphysician labor and capital, w is the wage rate for nonphysician labor, and c is the user cost of capital.[8]

[5] Some empirical justification of this assumption may be found in Donald Yett et al, where it was estimated that the elasticity of demand for *hospital* output with respect to *surgeons'* fees is 0.7.

[6] If the market for output is perfectly competitive, $|\partial Q/\partial P_T|$ will be infinite; otherwise, the individual hospital demand curve for output will have a negative slope.

[7] If the hospital received contributions, it may set a target loss equal to the contributions, but this will not alter our analysis. Moreover, after the fact the hospital may have a profit or loss, but this is assumed to result wholly from stochastic factors.

[8] The interpretation of the user cost of capital c is worth comment. When capital is provided through borrowed funds, the interpretation is clear; c is equal to $(r+d)P_K$, where r is the interest rate at which the funds were borrowed, d is the depreciation rate, and P_K is the price of capital goods. When unrestricted donations are used to pay for the marginal unit of capital, the user cost is $(r'+d)P_K$, where r' is the opportunity cost of using contributed funds for hospital physical capital, i.e., the rate which could have been earned on those funds if they had been invested elsewhere (say, in government bonds). When donations are made with the restriction that they be used for physical capital investment, they will affect the marginal user cost of capital only if the hospital receives so much in donations that it does not have to turn to any other source for funds for capital investment (unless, of course, the conditions for contribution of restricted funds specify a certain amount of the hospital's own funds as matching payments). If restricted donations fall short of the amount which, given the interest rate r, the hospital wishes to invest so that the hospital borrows, the relevant *marginal* user cost of capital must involve the interest and depreciation rates. Except for the case in which restricted donations are so large that the amount

The hospital is to be run so as to maximize the net incomes of the physicians on the staff at any point in time. If the number of physicians in the short-run analysis is taken as given at $\overline{M}$, the problem is to maximize $P_TQ - P_hQ$ (which is equal to $P_TQ - wL - cK$) subject to the production function (1), with the level of M set at $\overline{M}$, and the demand curve (2). This problem is obviously identical to that facing an orthodox profit-making firm with one input held constant. The marginal conditions for optimal employment of labor and capital are the same, namely, that marginal factor costs equal their respective marginal revenue or value products.[9]

It may be useful at this stage to contrast the model of the nonprofit hospital just developed with the orthodox model of the profit-maximizing firm. In the latter case, all labor inputs and capital services financed by debt are paid their competitive costs. Nondebt capital then obtains the residual income, which is usually assumed to consist of payment of the opportunity cost of that capital (normal profits) and economic profit. The only difference between this model and the physician-profit maximization model of the hospital is that in the latter it is the physician input, rather than the nondebt capital input, which obtains economic profits, the residual income. If a profit-maximizing firm submitted two bills for its services—one just covering the cost of labor and debt-financed capital, produced in a "nonprofit" firm, and the other from a separate legal institution covering the services of equity capital, the analogy would be complete.

II. Long-Run Individual Hospital Equilibrium

The number of physicians on the staff of any hospital obviously is not fixed, but is variable over time. What determines the size of the hospital's staff? The answer to this question depends critically on the assumption made about the hospital's staffing policy. We shall outline the results of three alternative policies—closed staff, open staff, and a policy in which new physicians can be hired by the hospital.

Closed staff

Many hospitals in the United States restrict staffing privileges: they do not permit any physician to join the hospital's staff just because he wishes to do so, even if he is licensed to practice medicine and surgery. The decision on whether or not to admit a new member to the staff (or whether to replace a member who has left) is made by the existing members of the hospital's staff of physicians. If we

of capital that is bought with them exceeds the amount which would be indicated by the marginal conditions, donations, whether unrestricted or restricted, are really equivalent to lump sum subsidies. Only if the hospital's price and output policies affect contributions (and within wide margins, they do not seem to) should contributions be treated as other than lump sum grants. In a world of uncertainty, however, the total of donations past and present may affect a lender's willingness to lend, since they provide collateral. Even in this case, current donations are likely to be important only if they are large relative to total nonborrowed capital.

[9] The model implies cost minimization in the sense that, given the physician input, quantities of labor and capital are chosen which, given their marginal supply prices, minimize costs. Cost minimization is also a characteristics of output-maximization models. However, normative conclusions that have been derived from empirical cost function studies regarding socially optimal scale of hospital facilities are considerably weakened when it is realized that there is no reason to suppose that, in comparisons across hospitals or over time, the physician input actually is constant, or even variable in a random way. The physician input has been left out of cost and production function studies of *U.S.* hospitals. Unless the physician input is specified or is known to be a constant ratio to K and L, there is no way of knowing the true social costs of all the inputs associated with any scale of output, and hence no way of determining the cost-minimizing scale. Observed decreasing hospital costs, for instance, may only represent a systematic increase in physician input with size. Furthermore, when we allow imperfect cooperation of the physician staff in our model in a later section, we will find that minimization of even nonphysician costs in the technical sense by the hospital is a very unlikely conclusion.

assume that once a physician is admitted to the staff, he has privileges identical to those of the existing members, the appropriate maximand for the hospital appears to be the maximization of net income per physician, Y_M. Physicians will be willing to add members to the staff as long as it causes each member's net income to rise.

This implies that the objective function to be maximized is

$$(4) \qquad Y_M = (P_T Q - cK - wL)/M$$

subject to the production function (1) and the demand curve (2). The necessary first-order conditions for an extremum become

$$(5a) \qquad w = P_T \frac{\partial Q}{\partial L} + \frac{\partial P_T}{\partial Q} \frac{\partial Q}{\partial L} Q$$

$$(5b) \qquad c = P_T \frac{\partial Q}{\partial K} + \frac{\partial P_T}{\partial Q} \frac{\partial Q}{\partial K} Q$$

$$(5c) \quad Y_M = P_T \frac{\partial Q}{\partial M} + \frac{\partial P_T}{\partial Q} \frac{\partial Q}{\partial M} Q$$

In long- or short-run equilibrium, the physician-hospital conglomerate firm that we have postulated will equate the marginal supply price of all nonphysician inputs to their respective marginal revenue or value products. However, in our model, physicians all share equally in the residual income of this health enterprise, the shares depending on their assumed equal shares of a total output. Condition (5c) states that physician staff size is determined in long-run equilibrium by equating the marginal revenue product of physicians to the net average revenue product of the physician staff. This makes intuitive sense. The hospital "pays" for new physicians by allowing them a proportionate share of total output and, hence, of net revenues. Staff physicians will want to welcome warmly a new member as long as his contribution to total revenues of all staff physicians is greater than the average current income per physician which he receives.

Of course, condition (5c) cannot be satisfied unless there are physicians willing to work at the hospital for the earnings available. That is, the equilibrium value of Y_M must be at least as large as the income stream available to a physician in his next best opportunity. There will be a supply curve of physicians to any hospital. The shape of this curve will depend in part on the income a physician could get in other hospitals, and his valuation of other uses of his time, both as leisure and as office practice.

Figure 1 depicts the long-run equilibrium position of a hospital operating in an urban, physician-intensive environment. The physician supply curve, SS, may therefore be assumed to be approximately infinitely elastic, and we also assume that it is at a low level relative to income possibilities in this particular hospital. Within the hospital, capital and nonphysician labor take on short-run optimal values as physician staff size, M, varies along the horizontal axis. ABC thus represents the maximum attainable income per physician for each specific value of M. Returns to scale and elastic demand lead initially to the upward sloping segment of ABC, but eventually decreasing returns and diminishing marginal revenue cause the curve to turn down. The maximum maximorum, Y_M^*, of this set of short-run maximums is reached at the intersection of the marginal revenue product and net average revenue product curve, when physician staff size reaches its long-run equilibrium value of M^*.

This model is very similar to those developed by Benjamin Ward (1958, 1970), Evsey Domar, Walter Oi and Elizabeth Clayton, Jaroslav Vanek, and others who explain the economic behavior of the Soviet collective farm or the Yugoslav producers' cooperative. The physician

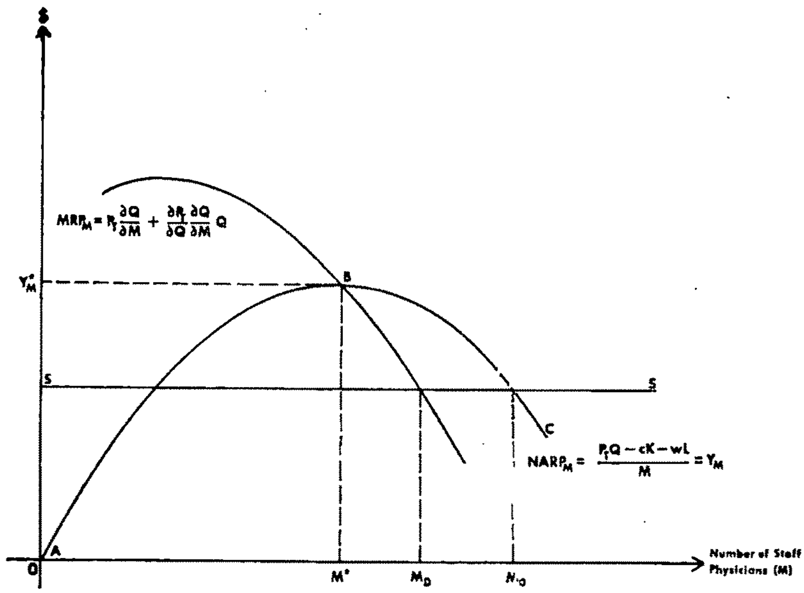


FIGURE 1

plays a role analogous to that of the Russian peasant and the Yugoslav worker. Institutionally, the arrangements differ because the physician receives his "share" of the enterprise's income directly from the sale of his output, whereas the worker on a collective receives it from a common pool. But this is because, in the case of the hospital, output can usually be directly assigned to particular staff members.

Our model of the hospital shares certain of the seemingly paradoxical conclusions of these cooperative-collective models. Supply response to changes in product and factor market conditions can be in perverse directions. An upward shift in the demand curve for hospital output could result[10] in a new equilibrium with higher price levels, *lower* output, and *fewer* physician staff members to share the greater total net revenues of the physician-hospital conglomerate enterprise. "Members have an incentive to contract membership to hoard the spoils" of a demand increase (see Oi and Clayton, p. 43). An increase in a factor price may lead to an expansion of operations to help spread the misery around among a larger group of individuals. On the other hand, a lump sum subsidy, such as a philanthropic contribution, will decrease output and staff size. "Even when the co-op moves in the same direction as a capitalist firm, its response is usually more sluggish. For market stability, the picture is not particularly reassuring" (see Domar, p. 739).

The reason for this result can be sketched out briefly:[11] Suppose a hospital faces a given price for output, and price increases by some amount. This will produce a proportionate increase in the marginal revenue product of physicians, assuming that K and L are held constant, but the income per physician (Y_M) will increase more than proportionately. The

[10] Unambiguous results concerning the direction of supply response cannot be determined unless specific restrictions are placed on the nature of the demand shift and the form of the production function.

[11] For a more extensive development, see the references cited in the preceding paragraph.

maximum average income would be attained at a smaller value of M.

The values of K and L will not, of course, remain constant. They will increase in response to increases in their marginal products. But M^* will still decline unless increases in K and L increase the marginal revenue product of physicians and do so by a large enough amount to offset the increase in average income. This need not happen; Q and M^* may decrease, or not increase much, whereas in a profit-seeking firm the use of an input and the amount of output would almost certainly increase with a rise in the price of output.

Discriminatory Sharing or Hiring Model

To consider the consequences of altering the "equal sharing" assumption, we suppose that the hospital depicted in Figure 1 is allowed to hire physicians at the supply price, OS. The hospital will then organize production in the same way as would a profit-maximizing firm. The physician input will be increased until the marginal revenue product and the marginal supply price of physicians' services are equal to each other. Staff physicians will be able to capture, in their own incomes, the excess of the marginal products of the infra-marginal hired physicians over their supply price. Equilibrium will be at M_D. But note that there is explicit discrimination in returns to homogeneous labor in this situation; the ability to sustain a stable equilibrium under these circumstances is highly suspect. As Pauly has shown, a system of clubs in which some identical persons receive less than others is not likely to be stable. It may well be that the internship-residency programs so prevalent in the United States may be an institutionalized method of getting around the inherent instability in the discriminatory hiring model by creating artificial, functionally viable distinctions among homogeneous physicians.[12] Determination of the ratio of "partner" physicians to "hired" physicians is likely to be arbitrary. The economic well-being of those physicians left with full staff privileges varies inversely with this ratio.

Open Staff Model

Alternatively, we can retain the "equal shares" assumption, and examine the economic behavior of hospitals that do not restrict entry to their physician staff. Any licensed physician who chooses to do so may become a "full partner" in the firm. Equilibrium at M_o in Figure 1 is characterized by equality of average income per member of the physician staff and the marginal supply price of physicians services.

Of course, for a hospital in a rural area with few physicians, or in an area where many physicians have attractive alternatives to membership on that hospital's staff, the situation might be somewhat different. Faced with a sharply rising supply curve for physicians' services, $S'S'$ in Figure 2, the physician staff will be in equilibrium at M_{oo}. The hospital might as well call its policy "open staffing," since it would be willing to add new members in order to move up the rising part of the net average product curve and increase income per staff member. Such a "frustrated closed staff" hospital only needs to adopt a closed staff policy when the number of staff members reaches M^*.

Note that the discriminatory-hiring and open-staff hospitals, either the "true"

[12] Interns and residents serving in *U.S.* hospitals currently comprise about 15 percent of all MD's and, of course, provide a considerably higher percentage of hospital based physicians' services. In several of the cost function studies, there have been attempts to estimate the effect these physician trainees have on hospital costs and revenues. Our model would suggest that it might prove more fruitful to analyze the ways in which trainees can increase the incomes of those physicians with regular staff privileges.

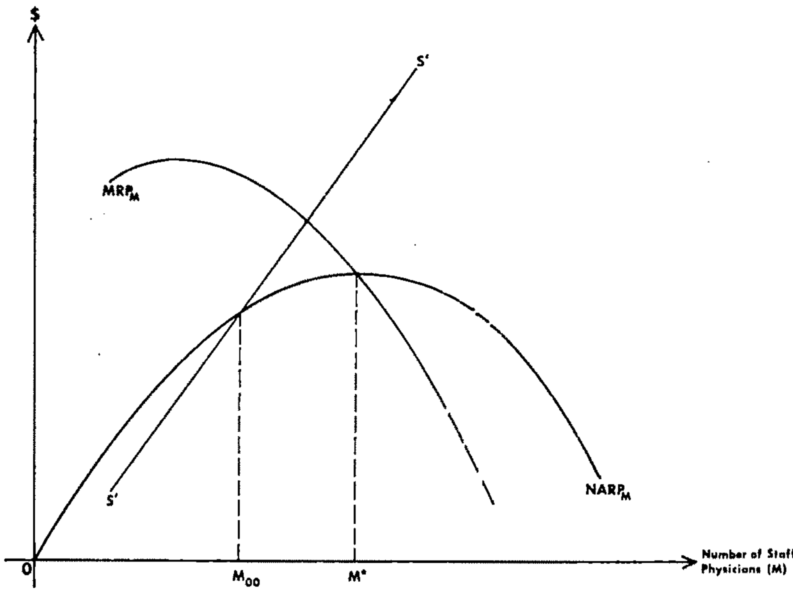


FIGURE 2

open-staff hospital or the frustrated closed-staff one, do not exhibit the potentially perverse supply responses that were uncovered in the closed staff model. Increases in demand will lead to increases in physician staff and output.

III. Long-Run Industry Equilibrium

The closed-staff hospital in the long run tries to adjust its physician staff size to achieve maximum income per physician. But if physicians in a hospital achieve incomes below those of other identical physicians on the staffs of other hospitals in that area or in other areas, the low-income physicians may wish to join the staffs of high physician-income hospitals. If they are prevented from joining the staffs of existing hospitals, as would occur if those hospitals were large enough so that average income was maximized, it will be worthwhile for M^* of those rejected physicians to join a new hospital, a duplicate in size of the old one. Indeed, the formation of new hospitals will continue as long as higher incomes can be earned, bidding up the supply price of physicians and bidding down the price of final output (and hence the curve of average income) until a position of long-run equilibrium is attained at which average income, marginal income, and marginal supply price are equal to each other and equal across hospitals. In the open-staff model, long-run equilibrium is also reached by the formation of new hospitals, but here the new hospitals draw off excess members from existing hospitals, raising average incomes since average size shrinks. This process will continue until formation of a new hospital does not raise average income. Finally, in the discriminatory-hiring model, physicians who are paid less than staff physicians (i.e., hired) will find it advantageous to try to join a hospital of their own in which they are all staff members and receive incomes equal to those of physicians in existing hospitals.

Thus these models have identical long-run industry equilibria (when the number of firms is variable), but the closed-staff model has a radically different long-run firm equilibrium, and may move in a perverse direction from that of the discriminatory-hiring firm. The open-staff model, on

the other hand, responds in the same way as would the discriminatory-hiring (or profit maximizing) hospital, but is likely to have a quantitatively different response. Although the final industry outcomes under these institutional arrangements are the same, the adjustment process by which that outcome is reached is very different in each case.

IV. Imperfect Cooperation on the Hospital Staff

In this section we wish to relax the assumption of perfect cooperation by the hospital's staff of physicians. The *individual* physician may in fact have direct control over the process of producing output. He is able to order use of nursing and other inputs for his patients, in his character of professional expert. At best, "the hospital" can determine only a stock of inputs; the physician controls the flow of services from them. This is in contrast to a producers' cooperative (or even a nonprofit university) in which, presumably, a central management is somewhat more able to prevent individualistic behavior. Moreover, each physician's income from the hospital-connected services depends entirely on his own output, in contrast to the producers' cooperative model in which an individual worker's income partly depends on his own output but also depends on his share of total profits or surplus income. The physician shares group income insofar as he shares group output.

We assume, as before, that the hospital price is set at a level which permits the hospital to break even. It is now easier to see why this assumption is itself a consequence of the physician profit-maximization hypothesis. Other things being equal, hospital profits mean higher hospital prices, which reduce physician incomes. To be sure, hospital profits could be used for capital investment which may enhance future physician incomes, but 1) physicians may not want to invest in the hospital—and 2), as Eirik Furubotn and Svetozar Pejovich have noted, the fact that property rights are not vested in investment from profits makes it even less desirable than owned investment.

Since prices for services are not (yet) set at marginal cost,[13] and since the physician does have the power to direct the application of labor and capital services in producing his output, noncooperative behavior can occur. Suppose the hospital follows the policy of charging an all-inclusive daily rate. Suppose ordering the use of more labor in the production of some output will permit the physician to charge more for that unit of output, because the "quality" of the output is enhanced. If the application of more nonphysician labor raises the price P_T^i that can be charged by the ith physician but if the same price P_H is charged for all units of hospital output, regardless of whether or not extra labor is ordered, the physician will want to use labor for his Q_i patients up to the point at which the following equality is satisfied:

$$\frac{\partial P_T^i}{\partial L} Q_i = \frac{Q_i}{Q} w \tag{6}$$

That is, the individual physician only considers a fraction Q_i/Q of the cost of the labor he orders to be employed, whereas group profit maximization would require him to take account of the costs his action imposes on other physicians' patients. Consequently, imperfect cooperation leads

[13] In fact, it does not appear that hospitals do price at marginal cost. They have moved away from the use of an all-inclusive daily rate, the same for all patients, toward itemized bills, but they are still far from charging marginal cost as a price. And in theory, it is unlikely that prices could be set equal to marginal cost. Unless marginal cost just equals average cost, marginal cost pricing by the hospital of its services would violate the no-profit, no-loss constraint assumed above. Moreover, the use of different prices for different qualities of outputs involves transactions and information costs, which may be substantial.

to the use of too much labor relative to the amount that would be used if group incomes were maximized with perfect cooperation.

How great these departures from perfectly cooperative usage will be depends upon how well the staff as a group can control the actions of its individual members. It seems reasonable to suppose that the magnitude of the departure from cooperative behavior should get larger as the size of the staff gets larger.[14] There are at least three reasons for making this assumption. 1) When the staff size is small, each physician bears a larger share of the cost of his own actions. 2) When the staff size is small, departures from cooperative behavior on the parts of others are more noticeable to any single member. 3) When the staff size is small, mutually agreeable group decisions are more likely to be arrived at, since the cost of decision making is less.

In terms of the geometric analysis, introduction of imperfect cooperation in this way shifts the curve of net revenue per member down and to the left. The conclusion is obvious; lack of perfect cooperation makes the (second best) optimal size of the staff and hospital smaller than it would otherwise be.

V. The Effect of Insurance

To this point we have assumed that patients confronted the full market price for care. Insurance coverage of medical costs may weaken the applicability of this assumption. The hospital is typically paid on the basis of costs incurred. If there are no copayments or deductibles at all, this arrangement effectively eliminates market control over the hospital component of the price of hospitalization, since higher hospital costs and prices have no effect on the cost actually paid by the patient (except through the premium, which only matters if the premium rises so high that the person drops coverage). Moreover, the allocation of costs within the hospital to insured patients, especially Blue Cross members, is often done on average cost basis. The only restraint on the physician's prescribing the use of hospital capital and labor for his fully insured patient is the upward pressure his behavior would exert on the prices paid by his noninsured patients. Conversely, under average cost allocation some of the market discipline is lost for the noninsured by the transfer of some costs to the insured.

If every patient had full coverage cost-based hospital insurance, there would be no constraint on the amount of capital and labor that physicians combine with their services. Capital and labor would therefore be employed up to the point at which the marginal contribution of each to the physician's revenue was zero. This produces "Cadillac-quality" medicine. The only constraint on the use of these inputs would be offered by the upper limit on the number of things a hospital can do for a patient which might have some justification. Over time, technological change might be expected to relax even this constraint.

When insurance covers part of each patients' hospital bill, the factor prices of hospital inputs K and L are effectively reduced as far as the group of physicians is concerned. One would expect an increase in the usage of hospital inputs relative to physician inputs for producing a given output. Hospital unit costs would rise. Thus our model provides an explanation of the positive relationship between hospital insurance coverage and hospital unit prices and costs found by M. Feldstein (1971). In addition, our model yields the important result that, *ceteris paribus*, increased

[14] A similar analysis of the effect of size is provided by Mancur Olson. Other applications to medical care can be found in the work of Newhouse (1972) and Mather Lindsay.

hospitalization insurance should increase physician prices and physician incomes.

VI. Toward a Theory of the Not-For-Profit Enterprise

To this point we have taken the not-for-profit nature of the typical American hospital as given. In this section we offer an explanation of why this organizational form has arisen.

There are two ways in which the hospital might operate if it were on a profit-maximizing basis. It might combine the services of nonphysician labor and physical capital, and sell them to or through the physician, who combines them with his own input to produce the output of hospitalization services. Alternatively, the hospital might perform the job of combination itself, hiring the physician input and selling the final output. In either case the direct control over the use of nonphysician labor and capital would not be held by the physician. In the first case, he would have to use the market for control, which is not always efficient in the sense of minimizing all costs, including transactions costs, as direct control. In the second case, he himself would be under the direction of the supplier of equity capital. It is surely possible that there are some products which are not produced efficiently when a representative of the owners of equity capital directs their production, or when the market is used to organize the production process instead of the use of direct controls within a single organization. The most efficient method might be for the supplier of an important component of the labor input to direct the production process.

This would tend to happen when human capital is important in the production of some output and when the flow of services of that human capital cannot well be directed from outside, but is controlled by the person in whom the capital is embodied. As Armen Alchian and Harold Demsetz have noted, the wage system tends to break down when marginal products cannot be monitored closely. This may be a reasonable conjecture in the case of the production of hospital services. Many of the decisions the physician has to make are decisions which cannot be supervised directly, and which have contingent outcomes. There probably needs to be some incentives for the physician. Financial interest in the outcome of his actions is one such incentive, and that incentive is at its greatest when the physician bears the full residual income, when the consequences of his actions are not spread over suppliers of physical capital.[15]

The production process requires some physical capital. In a labor-managed firm in which most of the assets of labor are embodied in nontransferable human capital, not all of the physical capital can be borrowed, since collateral cannot be provided. Another necessary condition therefore for the emergence of the not-for-profit form would seem to be the willingness of individuals to contribute for its equity capital. In principle, contributions could either be voluntary or provided through government. Where voluntary contributions are sufficient, government contributions would not be expected to emerge. On the other hand, voluntary contributions may arise in precisely those cases in which the government fails to act. They may also arise in cases in which the government through tax deductibility

[15] To see this, think of each physician as a "firm." The socially most efficient institutional arrangement is the one which maximizes the net present value of this firm. The present value is a contingency, depending on the state of nature (for example, what's really wrong with his patients, whether an epidemic occurs, etc.) and the amount of effort that the physician makes. The amount of effort, in turn, depends upon the share of profits the physician receives. The appropriate share for the physician, even given the greater risk he bears, may be approximately equal to one.

subsidizes private contributions. The source of contributions, whether unsubsidized voluntary, subsidized voluntary, or governmental, is not critical to the argument, except to the extent that one form (for example, government) implies more external control over physicians' actions than another.

These contributions could be motivated by a desire on the part of contributors to make output available to themselves or to those whom they would like to see consume it. That is, the motivation could either be based on the potential receipt of private benefits or of external benefits. Contributions are a logical way for potential purchasers of the outputs of labor-managed firms to make possible production of the output, which they or those about whom they are concerned will use. If there are barriers to entry by profit-seeking firms (as there are in higher education and, to some extent, in the hospital industry as well), potential consumers may be willing to contribute if that is the only way that output, which yields them consumers' surplus, can be made available. It is not surprising that private not-for-profit firms which sell output—hospitals, universities, symphony orchestras—tend to provide output which is used *not* by the poor, but partly by the contributors themselves.[16]

VII. Conclusion

The main thrust of the model we have suggested here, and the one which differentiates it from models of the not-for-profit hospital that have been suggested by others, is the use of the maximization of physicians' income as the characteristic function. The potential absence of perfect cooperation distinguishes it from similar models of producers' cooperatives. In a methodological sense, our model seems to be more attractive than those which simply assume that the not-for-profit organization maximizes a variable such as "quantity of output," because it explains what the organization does in terms of the economic motivation of those who control it.

More importantly, it appears to provide an appealing explanation of some peculiar characteristics of not-for-profit hospitals. The supposed quality consciousness of such hospitals, for example, is easily explained; "quality" is a synonym for application of nonphysician labor and capital in physician-income-enhancing ways, and noncooperative behavior could easily lead to "too high" quality. "Duplication of facilities" probably owes its existence to closed staffing and lack of perfect cooperation. Other aspects of hospital behavior could also be explained by considering their effect on physicians' income; the pattern of investment, for example, might be best explained by changes in the ability of capital to enhance physicians' incomes. The inelastic supply response of hospitals to Medicare and Medicaid is also consistent with our model.

Even the average size of hospitals, which seems, by most accounts, to be below the optimal or cost-minimizing level, can easily be explained. In the first place, empirically observed cost curves may be misleading, if we add the physician input. But more importantly, in a period of rising prices our model shows that hospitals will tend to be small, and for two reasons. First, smallness tends to permit maximization of net income per physician. Second, smallness is necessary to permit coordination of the medical staff.

[16] This last point is an important consideration. It is sometimes alleged that these firms have attained a nonprofit status so that they may better provide services to the poor. However, the recent experience in this country is for the poor to receive health services from government operated hospitals, to receive education in government operated institutions, and not to partake at all of the output of symphony orchestras, theatre groups, or private universities.

A narrower range of possible observations is consistent with our model than with the general utility-maximization model. Appropriate choice of the variables to enter the utility function can make almost any observed behavior consistent with utility maximization. In particular, the definition of quality is not clear. Our model specifies the variable in the objective function. In principle it will also predict quantitative as well as qualitative responses, in the sense that physician income can be measured while utility cannot. Unfortunately, at present hospital-specific data on physicians' incomes or prices do not exist which would permit us to provide a conclusive test of the model. Nevertheless, we hope that more data can be made available to test this model.

REFERENCES

A. Alchian and H. Demsetz, "Production, Information Costs, and Economic Organization," *Amer. Econ. Rev.*, Dec. 1972, *62*, 777–95.

W. J. Baumol and H. G. Bowen, "On the Performing Arts," *Amer. Econ. Rev. Proc.*, May 1965, *55*, 495–502.

J. Buchanan, "An Economic Theory of Clubs," *Economica*, Jan. 1966, *32*, 1–14.

E. Domar, "The Soviet Collective Farm as a Producer Cooperative," *Amer. Econ. Rev.*, Sept. 1966, *56*, 734–57.

M. Feldstein, *Economic Analysis for Health Service Efficiency*, Amsterdam 1967.

———, "Hospital Cost Inflation: A Study of Nonprofit Price Dynamics," *Amer. Econ. Rev.*, Dec. 1971, *61*, 853–72.

P. Feldstein, "Research on the Demand for Health Care," *Millbank Memorial Fund Quart.*, Part II, July 1966, 128–65.

E. Furubotn and S. Pejovich, "Tax Policy and Investment Decisions of the Yugoslav Firm," *Nat. Tax. J.*, Sept. 1970, *23*, 335–57.

P. Ginsburg, "Capital in Non-profit Hospitals," unpublished doctoral dissertation, Harvard Univ. 1970.

M. L. Lee, "A Conspicuous Production Theory of Hospital Behavior," *Southern Econ. J.*, July 1971, *38*, 48–59.

C. M. Lindsay, "Supply Response in the Financing of Medical Care," unpublished doctoral dissertation, Univ. Virginia 1968.

M. F. Long, "Efficient Use of Hospital," in *The Economics of Health and Medical Care* Ann Arbor 1964, 211–26.

J. Newhouse, "Toward a Theory of Nonprofit Institutions: An Economic Model of a Hospital," *Amer. Econ. Rev.*, Mar. 1970, *60*, 64–73.

———, "The Economics of Group Practice," *J. Hum. Resources*, forthcoming 1972.

W. Oi and E. Clayton, "A Peasant's View of a Soviet Collective Farm," *Amer. Econ. Rev.*, Mar. 1968, *58*, 37–59.

M. Olson, *The Logic of Collective Action*, Cambridge 1962.

M. Pauly, "Clubs and Cores," *Publ. Choice*, fall 1970, *9*, 53–65.

M. Reder, "Some Problems in the Economics of Hospitals," *Amer. Econ. Rev. Proc.*, May 1965, *55*, 472–80.

C. Stevens, "Hospital Market Efficiency: The Anatomy of the Supply Response," in H. Klarman, ed., *Empirical Studies in Health Economics*, Baltimore 1970.

J. Vanek, *The General Theory of Labor-Managed Market Economies*, Ithaca 1970.

B. Ward, "The Firm in Illyria: Market Syndicalism," *Amer. Econ. Rev.*, Sept. 1958, *48*, 566–89.

———, *The Social Economy*, New York 1970.

D. Yett, L. Drabek, M. Intriligator, and M. Kimball, "A Macroeconomic Model for Regional Health Planning, *Econ. and Bus. Bull.*, fall 1971, *24*, 1–21.

Capital, Distribution, and the Aggregate Production Function

By DONALD J. HARRIS*

Recent results in capital theory concerning the reswitching of techniques of production have shaken the foundations of a neoclassical parable according to which the total quantity of output per man is supposed to be a function of the total quantity of capital per man—the Surrogate Production Function—which can be used to "predict all behavior" in the sense of the wage and profit rates that would prevail in different long-run equilibria or steady states.[1] The source of the difficulty, it would appear, lies in the fact that the neoclassical parable attempts, as it were, to kill two birds with one stone, namely 1) to provide a general representation (or "surrogate") of realistic technologies involving production with heterogeneous commodities, and 2) to link the determination of the distribution of income directly to the technology itself and to the relative size of "factor endowments." It turns out that, in general, not only are realistic technologies not representable by such a function but also there is no one-to-one correspondence between techniques of production, the quantity of capital per man, and the distribution of income. These results have thus opened the way to a critique of the marginal-productivity theory of distribution in terms of its logical foundations (see P Garegnani). At the same time, they enable a return to certain classical conceptions concerning determination of income distribution in a capitalist economy (see A. Bhaduri (1969), Joan Robinson (1971), ch. 3).

The central proposition which is called into question by the recent results is that concerning the existence of a monotone-decreasing relation between the quantity of aggregate capital per man (the capital-labor ratio) and the rate of profit.[2] It has been shown that, in general, there exists no such relation: the relation between capital per man and the profit rate may be of any form. However, discussions concerning this relation have been unclear as to the exact features which account for its shape. Reswitching of techniques no doubt enters into it. But, even in the absence of reswitching of techniques, there may not be a one-to-one correspondence

* Associate professor of economics, University of Wisconsin, Madison. I have benefitted from discussions with Joan Robinson, Krishna Bharadwaj, Luigi Pasinetti, and John Eatwell and from correspondence with Luigi Spaventa.

[1] The concept of the Surrogate Production Function is due to Paul Samuelson (1962). The reswitching results are discussed in the 'Symposium on Paradoxes in Capital Theory." For a review of the issues and the controversy surrounding them, see G. C. Harcourt, Robinson (1970).

[2] This relation is derived from the production function through the requirement that the equilibrium rate of profit is equal to the marginal product of capital which is in turn a unique inverse function of the capital-labor ratio. On this, see the Appendix. Recognizing that there are "major troubles" with the neoclassical production function, J. R. Hicks, ch. 24, has proposed as an alternative a "sophisticated production function." He claims for this alternative that: "We can draw out a curve that expresses values of [capital per man] that are consistent with equilibrium for different [rates of profit]; it is highly probable that it will be a downward-sloping curve" (p. 298). The arguments considered here apply also to this conception. Another basic proposition of the neoclassical parable which is called into question by the recent results concerns the relation between consumption per man and the profit (interest) rate. Consideration of this is omitted here but the argument is easily extended to that relation.

between capital per man and the rate of profit.[3] Certain other features are relevant, some of which have customarily been discussed in terms of the category of "Wicksell Effects" (see Bhaduri (1966), Harcourt, pp. 397–400).

For the purpose of isolating these features, this paper seeks to identify more closely the various elements underlying the relationship between value of capital and rate of profit. The setting of the problem is in the comparison of different steady states which may be stationary or growing. There is no implication here of a *process of change* from one equilibrium to another. Different equilibria are to be regarded as if they were different planets with no connection between them. A clear distinction is maintained throughout between physical or technological relations and price relations. It is then possible to separate the differences involved as between one equilibrium and another into three categories: a *price effect*, a *composition effect*, and a *substitution effect*.[4] From this analysis, the proposition that the rate of profit is equal to the "marginal product of capital" also emerges in a new light.

I. Production, Prices, and Growth

It is helpful to carry out the analysis in the context of the familiar two-sector model of capital and consumption goods production.[5] Technology is specified in terms of a book of blueprints containing a list of the available techniques which may be finite or infinite in number. For any given technique, say α, the equilibrium price equations in terms of the consumption good as numeraire are:

$$(1)\quad p = b_1 w + v_1 p(\delta + r) \text{ (capital good)}$$

$$(2)\quad 1 = b_2 w + v_2 p(\delta + r) \text{ (consumption good)}$$

where

p = price of capital good
w = real wage rate
r = rate of profit
δ = rate of depreciation of capital good
b_1, b_2 = labor coefficients
v_1, v_2 = capital coefficients

$$\mu = \frac{v_2 b_1}{v_1 b_2} = \text{ratio of capital-labor coefficients.}$$

These yield equations for the *wage-profit curve* of the individual technique and the price of the capital-good which are, respectively,

$$(3)\qquad w = \frac{v_1(r^* - r)}{b_2 \xi(r)}$$

$$\xi(r) = \mu - v_1(\mu - 1)(r^* - r) > 0$$

$$0 \leqq w \leqq w^* = \frac{1 - v_1\delta}{b_2[1 + (\mu - 1)v_1\delta]}$$

$$0 \leqq r \leqq r^* = \frac{1}{v_1} - \delta$$

$$\frac{dw}{dr} < 0, \quad \frac{d^2w}{dr^2} \gtreqless 0 \quad \text{if} \quad \mu \gtreqless 1$$

$$(4)\quad p = \frac{b_1}{b_2 \xi(r)}; \quad \frac{dp}{dr} \gtreqless 0 \quad \text{if} \quad \mu \lesseqgtr 1$$

Similar conditions hold for techniques β, γ, . . . ω, allowing for differences in their respective production coefficients

There are as many wage-profit curves as there are alternative techniques in a given book of blueprints. Given any arbitrary

[3] An example of this was given by Pasinetti in his contribution to the "Symposium," p. 516. See also Spaventa (1968, p. 33).

[4] For a similar usage of the terminology of price and composition effects, see Spaventa (1968). A slightly different concept of "substitution-composition effects" is developed by Murray Brown.

[5] This model has the advantage that it is the simplest which captures a characteristic feature of realistic technologies, that is, production of heterogeneous commodities by means of commodities and labor. Its main limitation is that it does not allow for the existence of joint production in the use of fixed capital and for different time-patterns of production. Comparative equilibrium analysis in the context of this model has been treated by Hicks, Spaventa (1968, 1970), Garegnani, but with a rather different emphasis.

wage rate (profit rate), producers choose that technique which yields the highest profit rate (wage rate) or the wage-profit curve furthest out. The wage-profit curves corresponding to those techniques which are most profitable at some rate of profit generate an outer envelope or *wage-profit frontier*, curves that are dominated at every profit rate being discarded. With a finite number of discrete techniques, the frontier is made up of segments of the contributing curves and therefore has corners. Each corner corresponds to a *switch-point* between two intersecting techniques. With an infinite number of techniques the frontier is a smooth curve tangent at each point to one wage-profit curve. There may be *reswitching* of techniques such that a single curve contributes more than one segment or point to the frontier. In general, the frontier is negatively sloped and may have any curvature depending on the structure of the underlying technology. Because of the possibility of reswitching of techniques among those represented on the frontier, it is not possible in general to order techniques uniquely by the rate of profit.[6]

Add now the relations governing the levels of output (X_1, X_2), employment of labor (L), the stock of the capital good (M), the warranted growth-rate (g), and saving behavior. Saving behavior is represented by constant saving proportions s_r and s_w, respectively, for profits and wages, where $0 \leqq s_w < s_r \leqq 1$. As an alternative specification we consider the case of $s_w = s_r = s$. Accordingly, we have, for a given technique,

$$M = v_1 X_1 + v_2 X_2 \tag{5}$$

$$L = b_1 X + b_2 X_2 \tag{6}$$

$$X_1 = (g + \delta) M \tag{7}$$

$$pgM = s_r r p M + s_w w L \tag{8}$$

The model remains open with respect to one of the variables. To close it requires additional information which could be supplied by any of a number of possible theories. Without going into this, we focus instead on the characteristics of possible growth equilibria and on comparisons of such equilibria. Analysis of these properties provides us with the relations with which *any* relevant theory must be consistent.[7] As a basis for this analysis we obtain from the above system the following equilibrium relations:

$$c = \frac{X_2}{L} = \frac{v_1(g^* - g)}{b_2[\mu - v_1(\mu - 1)(g^* - g)]} \tag{9}$$

$$\frac{M}{L} = \frac{v_2}{b_2[\mu - v_1(\mu - 1)(g^* - g)]} \tag{10}$$

$$k = p\,\frac{M}{L} = \frac{\mu v_1 / b_2}{\xi(r)[\mu - v_1(\mu - 1)(g^* - g)]} \tag{11}$$

$$g = \frac{(s_r - s_w)\mu r + s_w r^* \xi(r)}{(1 - s_w)\mu + s_w \xi(r)} \tag{12}$$

[6] All of this is by now familiar, but bears repeating for the reason that it is basic to what follows. It may be noted also that everything said in this paragraph could as well be said about any technology in which commodities are produced by commodities and labor (compare Sraffa). The exception is that, with joint production (Sraffa) and with different time patterns of production (D. M. Nuti), the frontier may have both negative and positive slope. Otherwise, the simplifications introduced here are merely a matter of convenience. The wage-profit frontier is sometimes called a "Factor-Price Frontier" (see Samuelson (1962)). The present terminology is preferable because it does not presume a conception of profit as the price or reward of a factor of production. It is precisely this conception which is called into question by the results discussed here.

[7] This analysis is meaningful so long as the missing equation determines one of the variables in terms of exogenous parameters. A common formulation which fulfills this condition is the requirement that the system be on a steady-state path where the warranted growth rate is equal to the exogenous growth rate of labor. But there are others which would do just as well, for instance the Ricardian assumption of a wage rate equal to "subsistence," the Marxian assumption of a given rate of exploitation or Robinson's assumption of a rate of accumulation determined by "animal spirits."

where

$$0 \leqq g \leqq g^* = \frac{1}{v_1} - \delta$$

The accounting identities for net output per man (y) are

$$(13) \qquad y = w + rk = c + gk$$

II. Comparative Equilibrium Analysis: One Technique

In comparing different equilibria, special interest attaches to the relation between the rate of profit (r) and the value of capital per man (k). Assume, for simplicity, that there is only one technique which remains the same at every profit rate in the feasible range.

For the relation between k and r we have, by differentiation of (11),

$$(14) \qquad \frac{dk}{dr} = \frac{M}{L}\frac{dp}{dr} + p\frac{d(M/L)}{dg}\frac{dg}{dr}$$

It can be seen that there are two distinguishable components of the difference in k between different equilibria. The first is simply a difference in the price of the capital good associated with the difference in the profit rate. Call this a *price effect.* It is positive when the capital-labor coefficient is greater in the capital good sector and negative in the opposite case. When the capital-labor coefficient is the same in both sectors, the price effect is zero.[8] In this case, the classical labor-theory of value holds in its simplest form: price corresponds to labor value.[9]

[8] The reason for all this seems obvious when it is realized that there are two elements in cost: the wage rate and the profit rate, which vary in opposite directions. With a higher rate of profit (lower wage rate) costs rise more (fall less) in the sector which uses a higher proportion of the capital-good input to labor input. The relative price rises to reflect the difference in relative costs. If both sectors have equal capital-labor coefficients, the rise in one element of relative costs just offsets the fall in the other and consequently the relative price is unaffected.

[9] Setting $\mu = 1$ in (4) leads to $p = b_1/b_2$.

TABLE 1—COMPARISON OF EQUILIBRIA: SIGN OF dx/dr

x	$0<s_w<s_r$	$0=s_w<s_r$	$s_w=s_r=s$
$\mu>1$			
p	−	−	−
g	+	+	+
M/L	−	−	−
k	−	−	−
$\mu<1$			
p	+	+	+
g	?	+	−
M/L	?	+	−
k	?	+	?

The second component involves a more complex set of relations. With given saving proportions, when the profit rate is different the growth rate is likely to be different. This follows from the growth-profit relation (12) which represents the equilibrium condition for saving and investment. In addition, when the two commodities are produced by different methods ($\mu \neq 1$), a difference in the growth rate calls for a different stock of the capital good to produce the different composition of output which the difference in the growth rate entails (see equation (10)). Call this set of relations a *composition effect.*

The overall difference in the value of capital per man is the outcome of both the price and composition effects. The sign can be determined from the equations of the system.[10] These are shown in Table 1 for various cases. A question mark in-

[10] For the composition effect we have, from equations (10) and (12):

$$\frac{d(M/L)}{dg} = \frac{-v_2v_1(\mu-1)}{b_2[\mu - v_1(\mu-1)(g^*-g)]^2} \gtreqless 0 \text{ if } \mu \lesseqgtr 1,$$

$$\frac{dg}{dr} = \frac{\mu[(s_r-s_w)\mu + s_w(1-s_r)(\mu-1)v_1r^*]}{[(1-s_w)\mu + s_w\xi(r)]^2}$$

$$\gtreqless 0 \text{ if } \mu \gtreqless \mu^*$$

where

$$\mu^* = \frac{s_w(s_r-1)v_1r^*}{s_w(s_r-1)v_1r^* - (s_r-s_w)} < 1$$

dicates that the sign is ambiguous. This is because the price and composition effects are of opposite sign (when $\mu<1$, $s_w=s_r=s$) or because the composition effect itself is ambiguous (when $\mu<1$, $0<s_w<s_r$).

III. Many Techniques: The Discrete Case

When there is a finite number of techniques, the analysis in the previous section holds for each technique within the range of profit rates corresponding to the segment of its wage-profit curve included in the frontier. The additional feature which now concerns us relates to the comparison of equilibrium positions between which the technique of production is different.

We may call the difference in technique associated with a difference in the profit rate between two equilibria a *substitution effect*. The presence of the substitution effect means that the composition and price effects are no longer as sharply distinguishable as in the single-technique case. This is because the capital good may be different when the technique of production is different.[11] No quantitative expression in technical units of the capital good can then be given to the difference in the physical stock of each capital good associated with different equilibria. Neither can any meaningful comparison be made of the price of each capital good when they are different goods.[12] Instead, comparison has to be made of the value of capital expressed in terms of the numeraire commodity common to the different techniques.[13] This difference in value incorporates all three effects.

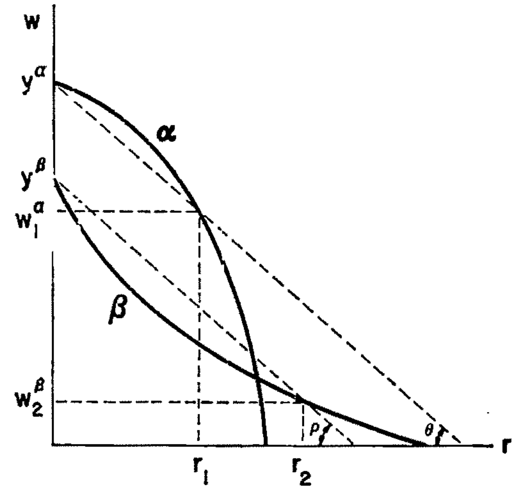


FIGURE 1

In comparing the various equilibria that are possible within the given conditions of technology and saving behavior, the same technique might be found in two different ranges of the profit rate. This reflects the existence of reswitching. The values of capital and net output per man for that technique in one range of the profit rate may be uniformly higher or lower than in the other range. These values may even be the same.[14] The result depends on the particular pattern of price and composition effects for that technique. There is, in this sense, no necessary inverse relation or one-to-one correspondence between "capital intensity" and the rate of profit.

But it is equally clear that there may exist no such correspondence *even in the absence of reswitching*. Some examples of

[11] Heterogeneity of capital goods must here be understood as a difference in their technical methods of production. This would be represented by a difference in the ratio of input proportions (μ) from one technique to another. Mere physical difference would imply as well a difference in units of measurement. It would be possible to conceive of a technology in which the capital good is the same for different techniques but we need not limit ourselves to such a special case.

[12] This particular limitation would not arise in a production system where each technique has many capital goods, for different techniques are then likely to have one or more capital goods in common.

[13] The consumption good is here taken as numeraire, but we could just as well have taken units of labor.

[14] For example, take the case of a technique with a straight-line wage-profit curve ($\mu=1$) which is tangent to the frontier over two discrete segments. For any two points, each on a different segment, both price and composition effects are zero so that the value of capital is the same.

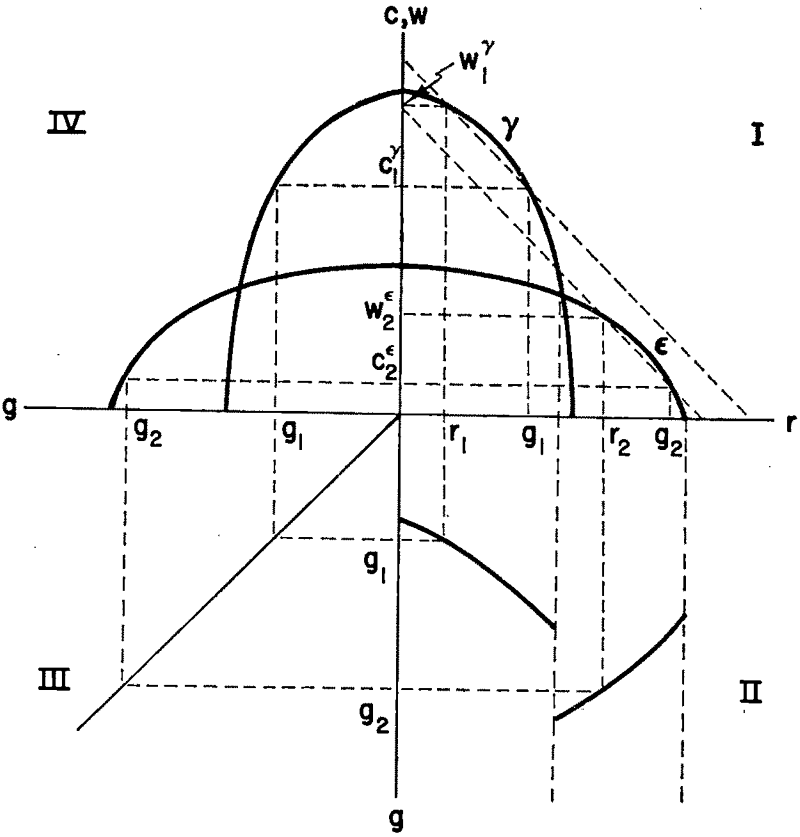


FIGURE 2

this are shown in Figures 1 and 2.[15] In Figure 1, assume a stationary state ($g=0$) so that $y=w^*$. At r_1, $k_1^\alpha=(y^\alpha-w_1^\alpha)/r_1 = \tan\theta$. At r_2, $k_2^\beta=(y^\beta-w_2^\beta)/r_2=\tan\rho$. By construction, $\tan\theta=\tan\rho$. Therefore $k^\alpha = k^\beta$. In Figure 2, assume saving behavior corresponding to the growth-profit relation shown in quadrant II. We know from the identities (13) that the value of capital per man satisfies $k=(c-w)/(r-g)$. At r_1, $k_1^\gamma=(c_1^\gamma-w_1^\gamma)/(r_1-g_1)$. At r_2, $k_2^\epsilon = (c_2^\epsilon-w_2^\epsilon)/(r_2-g_2)$. By construction, $k_1^\gamma = k_2^\epsilon$.

Two techniques may coexist, as at a switch-point. The associated equilibrium values can be identified for each technique separately using the common profit and wage rates. It will be found that the difference in the value of output per man between them will be just sufficient to yield profit at the given rate on the difference in the value of capital per man. This can be shown algebraically. By the income-accounting identity we have, for the two techniques, say α and β,

$$y^\alpha = w_0 + r_0k^\alpha, \qquad y^\beta = w_0 + r_0k^\beta$$

where w_0 and r_0 are the wage and profit rates at the switch-point. By subtraction we get

$$\frac{y^\alpha - y^\beta}{k^\alpha - k^\beta} = r_0$$

The left-hand-side of this equality is the ratio of the difference in the infinite stream

[15] The construction involved in Figure 1 is adapted from Garegnani and that in Figure 2 from Spaventa (1970).

of net output for each technique to the difference in the value of capital invested in each, output and capital being evaluated on a uniform basis in terms of the same consumption good and a common profit rate. Call this the *ratio of profitability* for the two techniques.[16] We may say in this case that the ratio of profitability is equal to the profit rate. This is a necessary implication of the condition that the two techniques are equally profitable at the given profit rate.[17]

IV. Infinite Number of Techniques

We can approach the case of an infinite number of techniques by conceptually arranging the techniques in the order in which they become eligible as r varies and by considering the coefficients of the techniques so arranged as functions of r. For this purpose we need only the coefficients v_1, b_2, μ, and the intercepts (w^*, r^*) of the corresponding wage-profit curves. Thus we write

$$(15)\quad v_1 = v_1(r); \quad b_2 = b_2(r); \quad \mu = \mu(r)$$
$$w^* = w^*(r); \quad r^* = r^*(r)$$

There is no a priori reason to expect these functions to have any particular form. Reswitching of techniques implies that the ordering is not unique. Particular restrictions can be placed on the technology so as to ensure a unique ordering. These must of necessity be arbitrary. One such condition which is sufficient but not necessary is that, for each successive technique represented on the frontier, w^* should fall and r^* rise as r rises.[18] Assuming $w^*(r)$ and $r^*(r)$ to be differentiable, this says that

$$(16)\quad \frac{dw^*}{dr} < 0, \quad \frac{dr^*}{dr} > 0, \quad \text{for all } r$$

Even with this restriction, there is no logical necessity for the coefficients v_1, b_2, and μ to vary in any particular way with r.[19] The restriction is satisfied in the special case when μ is the same *for all techniques*.

The equation of the envelope constituting the frontier can be found from the functions (15) and equation (3) for the wage-profit curve. Write this as

$$(17)\quad w = \phi[b_2(r), v_1(r), r^*(r), \mu(r), r]$$

The frontier represented by (17) is negatively sloped and may have any curvature. Thus, where differentiable,

$$(18)\quad \frac{d\phi}{dr} < 0, \quad \frac{d^2\phi}{dr^2} \gtrless 0, \quad \text{for all } r$$

[16] A ratio such as this could be computed for *any* arbitrary pair of techniques at *any* arbitrarily given profit rate. It can be computed for each pair of techniques in the entire book of blueprints from the *productivity curve* (see Appendix) constructed for that book at a given rate of profit and growth rate of the economy. In general, of course, the ratio of profitability would not be equal to the profit rate. A difference between this ratio and the profit rate is accounted for by the difference between the two techniques in the amount that would be left over from the net product for paying wages after deducting profits at the given rate on the respective values of capital. One technique is more (less) profitable than the other according as the ratio of profitability is greater (less) than the profit rate. Comparison of the profitability ratio for each successive pair of techniques with a predetermined rate of profit provides a basis for choosing the technique which is most profitable at that profit rate. When the ratio of profitability turns out to be equal to the profit rate, as at a switch-point, it is a matter of indifference which technique is chosen.

[17] Robert Solow (1967) has proposed that at a switch-point the rate of profit is also equal to the "social rate of return to saving." This proposition is further examined by Pasinetti (1969, 1970), Solow (1970).

[18] Compare Hicks, p. 154, and Michael Bruno et al. in "Symposium," p. 534. A related condition has been worked out by Brown in terms of variations in the composite coefficient v_2b_1 representing the indirect labor requirements per unit of consumption-good output.

[19] If it were to be assumed that δ is the same for all techniques, then $dr^*/dr>0$ implies $dv_1/dr<0$, so that the quantity of physical capital per unit of output in production of the capital good varies inversely with the profit rate. But no definite implication can be drawn from (16) for variation of b_2 and μ. In the special case of $\delta=0$ (the capital good lasts forever), $dw^*/dr<0$ implies $db_2/dr>0$, so that output per unit of labor in production of the consumption good varies inversely with the profit rate. Needless to say, these conditions tell us nothing about variation of the quantity of aggregate capital per man or of aggregate output per man in relation to the profit rate

Similarly, substitute the functions (15) into the growth-profit relation (12) and into equation (11) for the value of capital per man. The resulting relations may be written as

$$(19)\quad g = g[s_r, s_w, v_1(r), r^*(r), \mu(r), r]$$

$$(20)\quad k = k[b_2(r), v_1(r), r^*(r), \mu(r), g(\cdot), r]$$

The assumption of an infinite number of techniques does not necessarily imply that k and y become differentiable functions of r. However, for ease of exposition, it will be convenient to make the arbitrary assumption that these functions are continuously differentiable with respect to r. We can therefore characterize the difference between one equilibrium and another by taking the total derivatives as follows:[20]

$$(21)\quad \frac{dg}{dr} = \left[\frac{\partial g}{\partial v_1}\frac{dv_1}{dr} + \frac{\partial g}{\partial \mu}\frac{d\mu}{dr} + \frac{\partial g}{\partial r^*}\frac{dr^*}{dr}\right] + \frac{\partial g}{\partial r}$$

$$(22)\quad \frac{dk}{dr} = \left[\frac{\partial k}{\partial b_2}\frac{db_2}{dr} + \frac{\partial k}{\partial v_1}\frac{dv_1}{dr} + \frac{\partial k}{\partial \mu}\frac{d\mu}{dr} + \frac{\partial k}{\partial r^*}\frac{dr^*}{dr}\right] + \frac{\partial k}{\partial g}\frac{dg}{dr} + \frac{\partial k}{\partial r}$$

$$(23)\quad \frac{dy}{dk} = \left(\frac{d\phi}{dr} + k\right)\frac{dr}{dk} + r$$

Various components of the difference between equilibria appear from these relationships. Take the value of capital per man (k). With the technique held constant, the last two terms on the right-hand side of (22) represent, respectively, the composition effect and price effect. When the technique of production is different, there exists also a substitution effect. One might think of this as being represented by the bracketed terms in (21) and (22).[21]

Given the respective magnitudes for each equilibrium, the ratio (dy/dk) of the difference in value of net output per man to the difference in value of capital per man can be expressed in terms of the various components identified by equation (23).[22] An apparent similarity may be observed between this ratio and that which was called above the *ratio of profitability*. Nevertheless, the two should be clearly distinguished.[23] The ratio of profitability is computed for different techniques evaluated *at the same rate of profit*, reference being made either to a *switch-point* between two techniques (in which case the ratio is equal to the profit rate) or to a *productivity curve* representing all the available techniques. (See the Appendix.) It provides a way of representing the choice of technique through comparison of the ratio of profitability with the given rate of profit. In the present case, not only is the technique different, but the rate of profit is also different in the two situations compared. Moreover, this ratio is in general not equal to the profit rate. For what purpose can this ratio be used? The ratio in equation (23) has meaning only as a comparison of different growth equilibria or steady states. It can be used, so to speak, for accounting purposes, that is, to mea-

[20] It is assumed here, for simplicity, that different equilibria have the same saving proportions.

[21] When the capital good differs from one technique to another it makes no sense to compare their prices. It is therefore not economically meaningful to draw a strict separation between the difference in capital per man due to the difference in prices (the price effect) and that which is due to the technical difference in the capital good (the substitution effect).

[22] A similar formula was used by Bhaduri (1966) who called it the marginal product of capital. But, as we shall see, it can be regarded as the marginal product of capital only in a very special case.

[23] In the finite-techniques case this ratio is one of finite differences. It is expressed in terms of infinitesimal differences in the present case because of the assumption of differentiability.

sure the difference between equilibria. We may therefore call it the *accounting ratio.*

No general rule can be laid down concerning the sign and magnitude of these relations and therefore the direction of the overall difference between equilibria. The reason for this is obvious enough; it is impossible to obtain a unique form of the functions (15) that would be valid for all conceivable technical conditions. This remains so even when the sufficiency conditions (16) for absence of reswitching are assumed to hold. Particular results can be obtained under particular conditions of saving and/or technology. In the next section we consider two possible cases.

V. Some Special Cases

CASE 1:

$$s_w = 0, \quad 1 \geqq s_r > 0$$

In this case the growth-profit relation is independent of the technology. In particular, $g = s_r r$ for all r.[24] One technique may be associated with more than one profit rate due to reswitching. The variation of value of capital and net output in relation to the profit rate may take any form. It depends on the exact pattern of price, composition, and substitution effects.

If we take the special case of $s_r = 1$, this implies that $g = r$. We then get the following results.[25] By setting $g = r$ in (11), the value of capital per man at a given rate of profit becomes

$$k = \frac{\mu v_1/b_2}{\xi(r)^2} = -\frac{dw}{dr} \tag{24}$$

That is, it is equal to the slope of the wage-profit curve corresponding to the technique in use at that level of r. The same condition holds for every level of r, given the associated technique and its wage-profit curve. Since the wage-profit frontier is tangent to the wage-profit curve at every such point, their slopes are identical. It follows that, for all r

$$k = -\frac{d\phi}{dr} \tag{25}$$

Multiplying (25) by r/w leads to

$$\frac{rk}{w} = -\frac{d\phi}{dr}\frac{r}{\phi} \tag{26}$$

which defines the ratio of total profits to wages, or the distribution of net product, in terms of the elasticity of the wage-profit frontier. We also have from (25) and (18)

$$\frac{dk}{dr} = -\frac{d^2\phi}{dr^2} \gtrless 0 \tag{27}$$

which says that the value of capital may fall or rise with r depending on the curvature of the wage-profit frontier. Finally, by substituting (25) into (23) we get, for all r,

$$\frac{dy}{dk} = r \tag{28}$$

so that the accounting ratio is in this case equal to the rate of profit.

CASE 2:

$$\mu(r) = 1, \qquad \text{for all } r$$

Here the growth-profit relation becomes, for all r,

$$\begin{aligned} &g = (s_r - s_w)r + s_w r^*(r), \quad \text{if } 0 < s_w < s_r \\ &g = s r^*(r), \quad \text{if } s_w = s_r = s \end{aligned} \tag{29}$$

In either case of saving behavior, this relation is dependent on the technology through the function $r^*(r)$ which, in this case, represents the output-capital ratio

[24] Under the conditions of the Pasinetti theorem (1962), a similar relation holds without the special assumption that $s_w = 0$.

[25] Some implications of this case were pointed out by Bhaduri (1966), p. 288 and Spaventa (1970), p. 137: $g = r$ is a condition for the Golden Rule or neoclassical theorem to hold.

for the economy as a whole. Though in general nothing can be said about the form of $r^*(r)$, under the special conditions of this case we know that techniques can be uniquely ordered so that $dr^*/dr>0$ for all r. It follows that g is a continuous, monotone-increasing function of r. Thus there exists a unique relation between the growth rate, rate of profit, and technique of production over the entire range of available techniques.

By setting $\mu=1$ in (11), the value of capital per man at a given rate of profit is

$$k = \frac{v_1}{b_2} = - \frac{dw}{dr} \tag{30}$$

As in Case 1, it is defined by the technical coefficients representing the slope of the wage-profit curve corresponding to the technique in use. The results (25), (26), and (28) therefore hold also in this case. A special feature of this case is that the wage-profit curves of all techniques are straight lines. This implies that

$$k = - \frac{dw}{dr} = \frac{w^*}{r^*} \tag{31}$$

Since the condition (16) for unique ordering of techniques is fulfilled in this case, it follows that the value of capital per man is a monotone-decreasing function of the profit rate.[26]

The results of Case 2 can be readily explained. When $\mu=1$ there is no difference in the conditions of production of the two goods. Being technically the same, they can be regarded as one and the same good (except for units of measurement). Therefore their relative price is equal to unity (by appropriate choice of units) and is independent of the profit rate. This means that the price effect disappears. Furthermore, any composition of output can be produced with the same physical stock of the capital good. Thus the composition effect also disappears. There remains only a substitution effect which takes a special form. In particular, the consumption good is produced by itself and labor. All the available techniques can be ranked according to the increasing quantities of the consumption good required (as capital good) in production or according to the increasing quantity of net output of the consumption good produced. This ranking is unique and corresponds to the ordering of the techniques in relation to the profit rate. The "value" of capital per man, being equivalent to the physical stock of the consumption goods, also follows the same ordering. All of this constitutes what has come to be called the *one-commodity-model*.

[26] This case, which constitutes the basis for the Surrogate Production Function, is discussed by Garegnani (1970).

VII. The Accounting Ratio and the Profit Rate

The result that the accounting ratio (referring to two neighboring equilibria) equals the rate of profit is common to both Cases 1 and 2 and is similar to the result obtained at a switch point in the finite-techniques case. In Case 2, however, this result is susceptible of a further interpretation. Note, in particular, that in this case the difference in output per man and capital per man between two neighboring techniques consists entirely of a difference in their technical coefficients of production. The ratio of these differences (that is, the accounting ratio) is thus a technical datum independent of the rate of profit.[27]

[27] From this the possibility also arises of regarding the accounting ratio as representing the "marginal product" of the capital good itself because one equilibrium differs from another only by the addition of an infinitesimal quantity of the same capital good and by an associated increment in output of the same good. But when the capital good is different, and in the general case of many capital goods, no such interpretation is possible. For, then, the differences in capital per man and output per man can only be expressed as a value magnitude involving prices and the rate of profit.

Since it is also equal to the rate of profit it would appear, in comparing the two economies, *as if* a rate of profit existed for strictly technological reasons. Furthermore, since the profit share is equal to the elasticity of the wage-profit frontier, the distribution of income would appear to be uniquely related to the technology. These appearances would exist no matter what other explanation for the rate of profit and distribution of income might be offered by a theory relevant to the situations being compared. The technological reasons might even be made themselves to constitute such a theory. The point is, however, that those reasons and any such theory would have no applicability outside the framework of the special assumptions of this case and would therefore be, in general, of little or no interest whatsoever.

The fact that the same result (the accounting ratio equals the rate of profit) is found in Case 1 reflects the peculiar circumstances of saving behavior in that case and, beyond that, has none of the significance attached to Case 2. In particular, it does not mean that a higher value of capital is necessarily associated with a lower rate of profit. Neither is the accounting ratio susceptible of being interpreted as a technical datum which explains the rate of profit.

The common feature of both of these cases is that the value of capital is uniquely defined in terms of a technical datum.[28] The results of each case can be seen to follow directly from this condition. In the one case this condition is an accidental feature of a particular pattern of saving. In the other it is an accidental feature of a particular type of technology. Because, in Case 1, nothing further is implied concerning the technology, no general statement can be made about the relation between capital per man and the rate of profit. Because, in Case 2, a particular type of technology is assumed, there are particular implications for this relation: capital per man varies inversely with the profit rate.

There is no reason for assuming one or the other case as more likely to be found in reality than other conceivable possibilities. Indeed, in the present context, the conditions of the problem are sufficiently remote from reality to make any such statement meaningless. However, at the present level of abstraction, what we can see is that particular assumptions concerning the technology are *necessary* if there is to be an inverse relation between value of capital and rate of profit.[29] The restrictive nature of such assumptions and of the relation derived from them is evident from the fact that that relation no longer holds as soon as some slight change is made in those assumptions. Therefore, no theory based upon such foundations can lay claim to validity or generality or even to any significance as a "parable" or illustration of what exists in reality.

APPENDIX

The Productivity Curve and the Production Function

Let the profit rate be given at r_1. Then, for each technique there is a corresponding wage rate and relative price of the capital good in terms of the chosen numeraire. At a given growth rate of the system there is a corresponding physical stock of the capital good, a value of capital per man evaluated at the price corresponding to r_1 and an associated value of output per man. The values of capital and output per man obtained in this way for the entire book of blueprints form a *productivity curve*.[30] One such curve is shown in

[28] Note that, in either case, this feature is in no way related to the assumption of an infinite number of techniques. It could as well be found in a system with only one technique.

[29] In the neoclassical two-sector model (see Hirofumi Uzawa) the assumption that the consumption-good sector is the more "capital-intensive" serves the same purpose.

[30] This concept was introduced by Robinson (1956, p. 412) as an adaptation of one used by Wicksell.

Figure 3. The points α_1, β_1, γ_1, δ_1, ϵ_1, represent the values of capital and output for the available techniques at the profit rate r_1 equal to the slope of the ray r_1k. The *ratio of profitability* between techniques is the slope of the line joining two adjacent points.[31] Comparing the ratio of profitability for successive pairs of techniques with the given profit rate, we find that γ is the most profitable (it yields the highest wage rate Ow_1).

The same procedure could be carried out no matter how large the number of available techniques or the number of commodities.[32] The productivity curve (or its segments) would be positively sloped and display diminishing ratios of profitability between techniques going from left to right.[33] There is one point (or segment) to the left of which the ratio of profitability between techniques is greater than the given profit rate, to the right of which it is less. The technique(s) represented there is (are) the most profitable. The whole procedure is, in effect, a way of representing the choice of technique. It obviously does not provide a criterion for determining the profit rate for the latter is, so to say, *given in advance*.

At one profit rate (and corresponding set of prices) there is one productivity curve. At a different profit rate (say, r_2), the wage rate, relative prices of commodities and value of output and capital per man for each technique would in general be different. The productivity curve therefore has to be redrawn. Either the same or a different technique (or combination of techniques as in Figure 3) will be found to be most profitable on the new curve. The difference between the two equilibria can be analyzed in terms of the categories of price, composition and substitution effects developed in this paper.

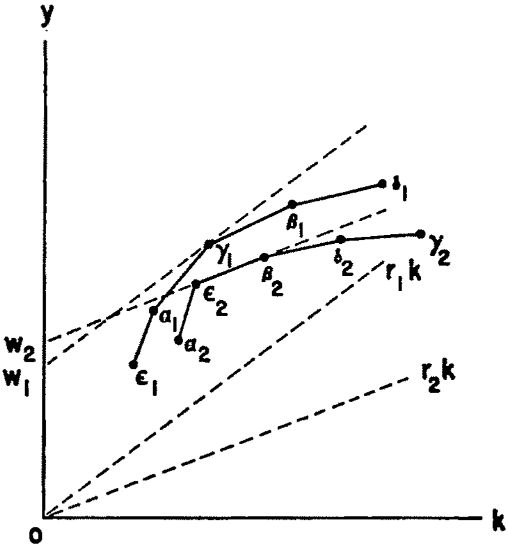


FIGURE 3

The production function relates output per man y to capital per man k, both measured in homogeneous units, such that

$$(A.1)\quad y = f(k);\quad f'(k) > 0,\quad f''(k) < 0$$

It is required that the rate of profit be equal to the marginal product of capital

$$(A.2)\qquad r = f'(k)$$

and from Euler's Theorem, we get

$$(A.3)\qquad w = f(k) - f'(k)k$$

To compare different equilibria, obtain from (A.1)–(A.3)

$$(A.4)\qquad \frac{dr}{dk} = f''(k) < 0$$

$$(A.5)\qquad \frac{dw}{dk} = -f''(k)k > 0$$

[31] The term "marginal product of investment" was used in this context by Robinson (1956, p. 412). In correspondence with the author she indicates that use of this term was misleading. On this, see Pasinetti (1969).

[32] Compare Michio Morishima, (1964), pp. 122–30. In the case of many consumption and capital goods, the relative prices would differ from one technique to another at the same rate of profit. To represent the values of output and capital for the different techniques on the same diagram the commodities must be evaluated at the same set of prices. The representation would then differ according to which set of prices is selected.

[33] This shape is a consequence of economic rationality, not of the technology. Any technique which, taken with another, had both a lower value of output and higher value of capital (that is, a negative ratio of profitability) would be irrelevant. Similarly, if between any three adjacent techniques the ratio of profitability were higher between the second and third than between the first and second, the second would be ruled out and comparison made directly between the first and third.

It follows from dividing (A.5) by (A.4) that

$$\frac{dw}{dr} = -k \tag{A.6}$$

As shown in the text, the wage-profit frontier is

$$w = \phi(r) \tag{A.7}$$

such that, where differentiable, $\phi'(r) < 0$ and $\phi''(r) \lesseqgtr 0$. (A.6) requires that

$$k = -\phi'(r) \tag{A.8}$$

Differentiating (A.8) and substituting into (A.4) leads to

$$\frac{dr}{dk} = -\frac{1}{\phi''(r)} = f''(k) < 0$$

which holds *if and only if* $\phi''(r) > 0$, that is, if the frontier is convex. If the frontier is linear or concave the production function cannot give the correct relation between k and r.

The particular case which meets the special requirements of (A.4) and (A.6) is Case 2 considered in the text, the essential feature of which is the assumption that a single homogeneous commodity (for example, corn) is consumed and used in production. The production function which describes the equilibrium relations that prevail in this case is a relation between homogeneous *physical* magnitudes. The capital or k in it is a physical quantity independent of prices and distribution. The productivity curve on the other hand, more generally relates quantities measured in *value* terms at the appropriate equilibrium prices and profit rate.

REFERENCES

A. Bhaduri, "The Concept of the Marginal Productivity of Capital and the Wicksell Effect," *Oxford Econ. Pap.*, Nov. 1966, *18*, 284–88.

———, "On the Significance of Recent Controversies on Capital Theory: A Marxian View," *Econ. J.*, Sept. 1969, *79*, 532–9.

M. Brown, "Substitution-Composition Effects, Capital-Intensity Uniqueness and Growth," *Econ. J.*, June 1969, *79*, 334–47.

M. Bruno, E. Burmeister, and E. Sheshinski, see "Symposium," pp. 526–54.

P. Garegnani, "Heterogeneous Capital, the Production Function and the Theory of Distribution," *Rev. Econ. Stud.*, July 1970, *37*, 407–36.

———, see "Symposium," pp. 555–67.

G. C. Harcourt, "Some Cambridge Controversies in the Theory of Capital," *J. Econ. Lit.*, June 1969, *7*, 369–405.

J. R. Hicks, *Capital and Growth*, Oxford 1965.

D. Levhari and P. A. Samuelson, see "Symposium," pp. 518–19.

M. Morishima, *Equilibrium Stability and Growth*, Oxford 1964.

———, see "Symposium," pp. 520–25.

D. M. Nuti, "Capitalism, Socialism and Steady Growth," *Econ. J.*, Mar. 1970, *80*, 32–57.

L. Pasinetti, "Rate of Profit and Income Distribution in Relation to the Rate of Economic Growth," *Rev. Econ. Stud.*, Oct. 1962, *29*, 267–79.

———, see "Symposium," pp. 503–18.

———, "Switches of Technique and the 'Rate of Return' in Capital Theory," *Econ. J.*, Sept. 1969, *79*, 428–31.

———, "Again on Capital Theory and Solow's 'Rate of Return'," *Econ. J.*, June 1970, *80*, 428–31.

J. Robinson, *The Accumulation of Capital*, London 1956.

———, "Capital Theory Up To Date," *Can. J. Econ.*, May 1970, *3*, 309–17.

———, *Economic Heresies*, London 1971.

P. Samuelson, "Parable and Realism in Capital Theory: The Surrogate Production Function," *Rev. Econ. Stud.*, June 1962, *29*, 193–206.

———, see "Symposium," pp. 568–83.

R. M. Solow, "The Interest Rate and Transition Between Techniques," in C. H. Feinstein, ed., *Socialism, Capitalism, and Economic Growth*, Cambridge 1967, 30–9.

———, "On the Rate of Return: Reply to Pasinetti," *Econ. J.*, June 1970, *80*, 423–8.

L. Spaventa, "Realism Without Parables in Capital Theory," *Recherches Recentes sur la*

Fonction de Production, Centre D'Etudes et de Recherches Universitaire de Namur, 1968, 15–45.

———, "Rate of Profit, Rate of Growth, and Capital Intensity in a Simple Production Model," *Oxford Econ. Papers,* July 1970, *22,* 129–47.

P. Sraffa, *Production of Commodities by Means of Commodities,* Cambridge, 1960.

H. Uzawa, "On a Two-Sector Model of Economic Growth," *Rev. Econ. Stud.,* Oct. 1961, *29,* 40–47.

"Symposium: Paradoxes in Capital Theory," *Quart. J. Econ.,* Nov. 1966, *80,* 503–83.

The U.S. Phillips Curve and International Unemployment Rate Differentials

By ROBERT J. FLANAGAN*

For much of the postwar period, the persistence of higher unemployment rates in the United States relative to other industrialized nations has perplexed economists and laymen alike. The earliest discussions of the problem emphasized technical differences in the measurement of the relevant statistical series, but work in the early 1960's showed that international unemployment rate differentials remained after corrections for differences in the statistical concepts and measurement (see President's Committee to Appraise Employment and Unemployment Statistics). Later studies emphasized the relative weakness of aggregate demand in the United States during most of the postwar period. The unsettling fact remained that the unemployment rates attained during the Korean War and other periods of strong demand pressure in the United States were higher than those observed abroad. A related explanation recognized the conflict of policy goals implicit in the Phillips curve and suggested that the officials responsible for monetary and fiscal policy in the United States preferred low rates of inflation by European standards. Implicit in this view was a common Phillips curve for each country and the notion that standard monetary and fiscal policies could bring the United States to European unemployment rates with no greater cost in terms of inflation. However, later evidence indicated that the tradeoffs between prices, wages, and unemployment rates differed between countries, so that the observed unemployment differentials were partially a matter of choice among conflicting alternatives, but partially the result of behavioral or structural differences which affect the position of the Phillips curve and the price change-unemployment tradeoff (see Ronald Bodkin et al.).

This paper will determine some of the main sources of the relatively "poor" Phillips curve tradeoff observed in the United States. The following sections derive some factors which may be responsible for differences in the slope and position of the curve and develop empirical estimates of the relative importance of these factors in the United States, Great Britain, and Sweden. The comparison is limited to these countries because of the difficulties in obtaining data which are sufficiently comparable to permit the conclusion that one is observing real differences rather than errors in measurement.

I Theoretical Determinants

In this section, a theory of the Phillips curve is outlined to derive parameters which govern differences in the slope and position of the Phillips curve over time and among countries. The theory generally follows the disequilibrium approach of Richard Lipsey and Bent Hansen, al-

* Assistant professor of labor economics, Graduate School of Business, University of Chicago. I wish to thank Bent Hansen, Lee Soderstrom, Lloyd Ulman, the managing editor, and an anonymous referee for useful criticism and suggestions on earlier drafts. This paper was prepared under Grant No. 91-05-68-70 from the Manpower Administration, U. S. Department of Labor, under the authority of title I of the Manpower Development and Training Act of 1962 as amended, but points of view which are expressed do not necessarily represent the official position or policy of the Department of Labor.

though comparisons to more recent neoclassical formulations are made where appropriate.

The familiar dynamic hypothesis underlying the present approach is that the rate of change of wages is proportional to the magnitude of the labor market disequilibrium, defined as the rate of excess demand. Thus

$$(1) \qquad \Delta \frac{W}{P} = k\left(\frac{D - S}{S}\right)$$

where $\Delta(W/P)$ is the time rate of change of real wages, D and S are the demand and supply of labor, respectively, and k is the reaction coefficient denoting the flexibility of wages in response to a given disequilibrium in the labor market. The appropriate measure of $[(D-S)/S]$ in the labor market is the job vacancy rate, V, minus the unemployment rate, U, where vacancies are defined as the difference between the demand curve (desired employment) and actual employment, and unemployment as the labor supply minus actual employment. However, in most empirical work a relation between the vacancy rate and the unemployment rate is postulated, and the unemployment rate is used as a proxy for excess demand. Movements along the unemployment-vacancies (UV) curve correspond to cyclical changes in the rate of excess demand, while shifts in the position of the curve correspond to shifts in the degree of structural maladjustment in the labor market.[1] Both k, the coefficient indicating the flexibility of wages in response to labor market disequilibrium, and the degree of "structural maladjustment" in the labor force will influence the position of the Phillips curve. The interpretation of structural maladjustment is elaborated in the following sections.

1. *Turnover*

Given the definitions of unemployment and vacancies, the parameters of the UV relation reflect the speed and extent of adjustment of actual to potential employment and actual to potential labor supply when the labor market is in disequilibrium. The study of this process is facilitated by disaggregating the stocks of unemployment and job vacancies into the flow per unit of time and average duration of each variable. Thus, the job vacancy and unemployment rates can be defined

$$(2) \qquad \begin{aligned} V &\equiv I_v D_v \\ U &\equiv I_u D_u \end{aligned}$$

i.e., as the product of the weekly incidence of new unfilled vacancies (unemployment) as a percentage of the labor force and the average number of weeks such vacancies remain unfilled (individuals remain unemployed). The identities define a steady state in which the weekly flows of new vacancies and new unemployment exactly match and replace the vacancies and unemployment generated (on average) D_v and D_u weeks earlier. The distinction is important in part because the determinants of the flow and duration of each stock are sufficiently diverse that the appropriate policies to reduce an observed stock of unemployment are not obvious a priori.[2]

[1] Prior to the appearance of the first Phillips curve, this relation was postulated and studied by J. C. R. Dow and L. A. Dicks-Mireaux who argued that frictional vacancies (at very low levels of demand) and frictional unemployment (at high levels of demand) will generate a non-linear relation. Lipsey applied a similar argument, and Edmund Phelps derives the same form from an argument based on labor turnover behavior. Jacob Mincer accepted the hyperbolic form and pointed out that movements along the UV curve coincide with movements along the Phillips curve. Empirical studies in Great Britain, Sweden, Australia, and the Netherlands (using data which are subject to the qualifications mentioned in the text) have uncovered highly significant non-linear relations between the aggregate level of vacancies and unemployment (see references in Hansen and Phelps).

[2] The distinction is also useful because it highlights the possibility that in tight labor markets an accelera-

When desired employment exceeds actual employment, the rate of incidence of vacancies is defined as

$$(3) \quad I_v = q(Q) + y(Q) + r + \Delta E^*(Q)$$

in which q is the quit rate, y the layoff rate, r the retirement rate, ΔE^* is the change in desired (cost minimizing) employment as a percent of the employed work force, and Q is the rate of excess demand in the product market. The rate of incidence of unemployment reflects the combined influence of quits, layoffs, and labor force entries, or

$$(4) \quad I_u = \alpha q(Q) + \beta y(Q) + \gamma p(Q)$$

in which p is the rate of gross labor force entrance, and α, β, and γ are parameters representing the propensity of each source of turnover to incur unemployment. From these definitions it can be shown that the disequilibrium behavior of the flows of new vacancies and new unemployment is not predictable in advance.

$$(5) \quad dI_v/dQ = q' + y' + \Delta E^{*\prime} \gtreqless 0$$

where q', $\Delta E^{*\prime} > 0$ $y' < 0$

$$(6) \quad dI_u/dQ = \alpha q' + \beta y' + \gamma p' \gtreqless 0$$

where $p' > 0$

The specified signs reflect considerable empirical evidence that voluntary frictions are positively related to excess demand in the product market while friction motivated by corporate maximizing decisions is negatively related to excess demand in the product market. It is clear that the disequilibrium behavior of I_v and I_u is theoretically ambiguous because of the opposing cyclical behavior of the underlying turnover flows. The results will depend on the relative strength of the (positive) cyclical elasticity of quits and entrants on the one hand and the (negative) cyclical elasticity of layoffs on the other, as well as the proportion of quits, layoffs, and labor force entrants which incur some unemployment.

2. *Duration of Unemployment and Vacancies*

Most of the cyclical variance in unemployment and vacancy rates is due to cyclical variations in the average duration of unemployment and vacancies D_u and D_v. However, there is some theoretical controversy concerning the proper interpretation of the observed phenomena. On the one hand, the cyclical sensitivity of D_u can be treated as an index of the severity of "structural" imbalances in the labor market which prevent the matching of workers and jobs. On the other, the duration of unemployment or vacancies (and thus the unemployment and vacancy rates) has been treated as search-related and hence voluntary.[3] Both formulations yield the same predictions concerning the cyclical behavior of longer term unemployment and vacancy rates. Therefore,

$$(7) \quad D_v = h(Q) \qquad h' > 0$$

$$(8) \quad D_u = j(Q) \qquad j' < 0$$

Thus the slope of the UV relation depends critically on the cyclical sensitivity of quits, layoffs, and labor force entrants, the proportion of each turnover flow which incurs unemployment, and the cyclical behavior of the duration of vacancies and unemployment. These same variables influence the slope and curvature of the Phillips curve, for as other writers have

tion in job-changing and labor force entry could increase the flow of new unemployment sufficiently to offset the effects of the falling average duration on the total unemployment rate, thus inducing a positive slope in the UV relation and the Phillips curve (see Bernard Corry and David Laidler). The empirical importance of this possibilty is assessed below.

[3] The equilibrium duration of unemployment (optimal duration of search) comes out of the theory of information as developed by George Stigler. More recently the theory has been recast to generate cyclical variations in D_u and D_v (see papers in Phelps et al.).

pointed out, the steeper or more curved the UV relation, the greater the change in excess demand $(V-U)$ and hence $\Delta(W/P)$ associated with a one unit change in the unemployment rate (see Hansen, Lipsey). Moreover, as indicated by the constant term, the position of the UV curve is also a function of the general level of turnover in the economy. Therefore, both the level and cyclical fluctuation in the turnover flows can be an important determinant of the slope and position of the Phillips curve and a fortiori an important source of international unemployment differentials.

3. *Expectations*

Although the theory of labor disequilibrium traced out in earlier sections is a theory of dynamic adjustments in real wages, the dependent variable in most Phillips curve studies is the rate of change of money wages. It is therefore necessary to include a variable which captures reactions to anticipated price increases.[4] Economists who emphasize the role of expectations in the wage determination process conclude that while a stable disequilibrium wage adjustment relation may be observed in the short run, there is no long-run tradeoff available to the economy, for the system will move back to an "equilibrium" rate of unemployment via one of several postulated mechanisms.[5] From this theory it follows that an unemployment rate which is lower than the equilibrium rate can be maintained by the application of monetary and fiscal policies only at the cost of an ever-accelerating rate of inflation as expectations and hence wage demands are constantly revised on the basis of price changes in current and previous periods.[6]

This argument has definite testable implications for the position of the *short-run* Phillips curve which have been stated clearly in this *Review* by Friedman:

> Stated in terms of the rate of change of nominal wages, the Phillips Curve can be expected to be reasonably stable and well defined for any period for which the *average* rate of change of prices, and hence the anticipated rate, has been relatively stable. For such periods, nominal wages and "real" wages move together. Curves computed for different periods or different countries for each of which this condition has been satisfied will differ in level, the level of the curve depending on what the average rate of price change was. The higher the average rate of price change, the higher will tend to be the level of the curve. [p. 9, fn. 5]

Since expected price increases are postulated to exert an independent influence on wage changes, the wage equation must be amended to include a separate expectations variable.

II. Aggregate Wage Changes in Sweden, Great Britain, and the United States

The objective of this section is to estimate aggregate wage change equations in each of the sample countries, to document differences in the slope and position of the Phillips curves for these three countries, and to offer some tentative judgements on the influence of price expectations and trade unions on the position of Phillips curves. The equations for each country are estimated on comparable quarterly data for the period 1951: II through 1968: IV by ordinary least squares. The dependent variable for each country is the rate of

[4] Unanticipated price increases will work through their influence on the rate of excess demand variable.

[5] The theoretical arguments and alternative methods of adjustment are specified in Milton Friedman and the papers in Phelps et al. In his introduction to the latter reference, Phelps admits the possibility of a long-run tradeoff which is nevertheless steeper than observed short-run relations.

[6] Others have argued that the maintenance of high demand will work out structural rigidities in the economy and that rising prices resulting from lower than equilibrium unemployment rates may erode the effect of institutional and legal constraints which impede employment, thus tending to reduce the equilibrium rate.

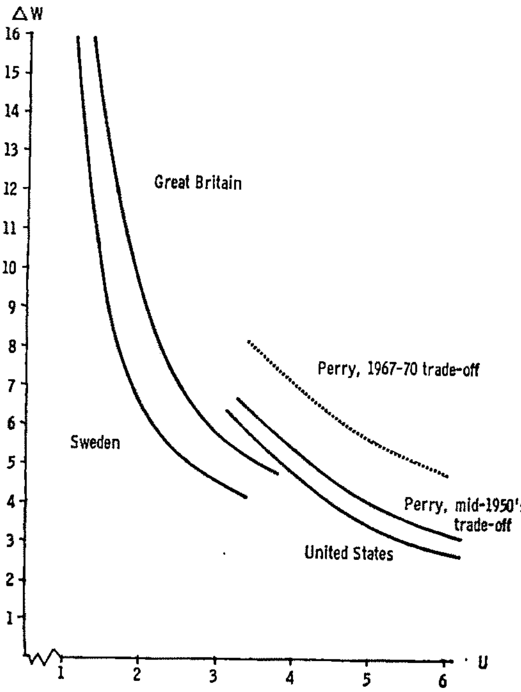


FIGURE 1. ESTIMATED PHILLIPS CURVES SWEDEN, GREAT BRITAIN. UNITED STATES, 1951–68

change of gross hourly earnings, which normally increases more rapidly than the rate of change of negotiated hourly wage rates because of the influence of overtime payments and wage drift.[7] Moreover, the unemployment rate in each country is adjusted to the concepts and measurement of the *U.S.* sample survey series.[8] The rate of change of the Consumer Price Index is used in the current period, lagged one quarter, and in an Almon distributed lag to capture the effect of anticipated price changes.[9]

In the subsequent regressions, ΔW is the four-quarter rate of change of earnings, ΔP is the four-quarter rate of change in the Consumer Price Index, and U is the average unemployment rate over the past four quarters. This form, which centers U about $1\frac{1}{2}$ months behind the ΔW variable, was found to perform best. The term Z is the sum of the coefficients of a distributed lag of past price changes and is included as an alternative to ΔP to test the proposition that workers obtain full compensation for price increases, but with a lag.[10]

The wage adjustment equations which were selected after experiments with alternative lag structures appear in Table 1, and the tradeoffs which these regressions imply are drawn in Figure 1.[11] For the

[7] The sectors covered by the earnings data differ between countries: the *U.S.* data are for the manufacturing sector, the Swedish data are for mining and manufacturing, and the British data are for the above plus construction, utilities, transportation, communication, and public administration. Sources are U.S. Bureau of Labor Statistics, Ministry of Labour *Gazette*, and *SCB, Allmän Månadsstatistik*. The British data are from a semiannual series which has been interpolated for intervening quarters.

[8] Unemployment data differ largely because many European series are by-products of unemployment insurance or employment service programs and tend to understate unemployment and the unemployment rate relative to sample survey data. A full analysis of these differences is included in the President's Committee to Appraise Employment and Unemployment Statistics. Since late 1961 a Swedish quarterly sample survey has provided unemployment rate estimates which closely correspond to *U.S.* concepts and measurement (see *SCB*, 1969). Estimates for earlier years were developed by linking the survey series to insurance and trade union series. For Great Britain a comparable unemployment rate series was developed by using Census and special survey materials to estimate the undercount of women and other groups in the official registration series (see U. K. Ministry of Labour *Gazette*). Details on the adjustments are in Flanagan, Appendix, and may be obtained from the author.

[9] Since indirect taxation has been an important component of Swedish fiscal policy in the 1960's, the Swedish regressions were tried with two price series—the rate of change of the total consumer price index and the rate of change of the net price index (see *SCB*, 1968). The latter excludes the effects of indirect taxes and differs from the former index only since 1959. There was no significant difference between the coefficients obtained from the alternative series.

[10] The reported results, selected after considerable experimentation, are for polynomials of degree 4, 3, and 5 fit to lags of 8, 5, and 12 quarters for the United States, Great Britain, and Sweden, respectively.

[11] The tradeoffs were generated under steady-state assumptions by substituting a price change equation (in which the rate of change of hourly earnings ΔW and import prices ΔI are related to the rate of inflation) for each country into the wage change equation with a single period rate of price change. The price change equations are:

TABLE 1—QUARTERLY EARNINGS CHANGE REGRESSIONS, SWEDEN, GREAT BRITAIN, UNITED STATES, 1951: II—1960: IV[a]

Country	Regression No.	Constant	Regression Coefficients						R^2	d
			U^{-1}	U^{-2}	ΔU	ΔP	ΔP_{t-1}	Z		
Sweden	1.	2.04		7.81 (2.03)[b]			.80 (.09)[b]		.78	1.08
	2.	3.39		3.27 (1.72)				.96	.82	.86
Great Britain	3.	2.75		15.87 (3.71)[b]		.41 (.08)[b]			.50	.86
	4.	3.01		12.59 (3.79)[b]	−3.23 (1.27)[b]	.48 (.08)[b]			.54	1.03
	5.	3.30		15.24 (3.67)[b]				.29	.52	.89
United States	6.	−.59	17.21 (2.01)[b]			.41 (.07)[b]			.72	.75
	7.	−.63	17.71 (2.28)[b]					.37	.75	.73
	8.	1.36		35.03 (4.09)[b]		.42 (.07)[b]			.72	.78

Sources: U.S. Department of Labor; U.S. Bureau of Labor Statistics; Ministry of Labour *Gazette; SCB Allmän Månadsstatistik; SCB*, 1969; Flanagan, Appendix.
[a] Standard errors are in parentheses and d is the Durbin-Watson statistic.
[b] Significant at 1 percent level.

United States, wage equations were estimated alternately using the U^{-1} and U^{-2} forms of the unemployment variable. Although there were no statistical grounds for favoring either specification over the other, the equation using U^{-1} is used to construct the tradeoff in Figure 1 to facilitate comparison with a recent study by George Perry. The tradeoffs derived by Perry, using somewhat different definitions of W and U, are included in Figure 1 and discussed below. On the other hand, the wage equation using the U^{-2} specification was required to conduct a test for structural independence of the wage equations for each country and is reported in regression 8. When a Chow test was applied, the computed F-ratios were above the critical value for each pair of countries, and the hypothesis that the observed differences in wage adjustment relations are the result of sampling variations was rejected (see Gregory Chow).[12]

Given the structural independence of the estimated Phillips curves, the regressions and the curves they generate appear to justify the following conclusions:

1) Over the range of observations provided by the 1951–68 period, the United

Sweden: $\Delta P=.77+.36\ \Delta W+.23\ \Delta I$
Great Britain: $\Delta P=-.12+.58\ \Delta W+.12\ \Delta I$
United States: $\Delta P=-.33+.58\ \Delta W+.12\ \Delta I$

All coefficients are significant at the 1 percent level. Alternative assumptions about the value of ΔI had a miniscule impact on the tradeoffs.

[12] The computed F-ratios are 10.15 for the United States and Sweden, 7.04 for the United States and Great Britain, and 7.49 for Great Britain and Sweden. The critical value was 3.82.

States experienced the least favorable Phillips curve. However, if one extrapolates the curves for the European countries, it appears that at unemployment rates over 4 percent, Great Britain and eventually Sweden would have experienced more wage inflation than the United States. This reflects the relatively high rate of autonomous wage increases in the two European nations as noted below. The conclusion of a relatively "inferior" *U.S.* tradeoff at lower unemployment rates is strongly reinforced by a comparison with the Phillips curve recently derived by Perry for the mid-1950's. For most of the period covered by the present study, Perry's estimates, which weight each member of the labor force by the relative hours and income received in the relevant age-sex cell, imply that a given ΔW is associated with about one-half of 1 percentage point more unemployment than the Phillips curve derived from regression 6. Moreover, for the period since 1967, Perry's work indicates that a shift in the composition of the labor force toward high unemployment groups has shifted the *U.S.* Phillips curve still further from European experience over the full range of postwar unemployment rates. The evidence supports the conclusion that the relatively high postwar unemployment rates in the United States reflected a less favorable tradeoff available to monetary and fiscal policy as well as the maintenance of a greater gap between actual and potential output.

2) Autonomous wage changes, as reflected in the constant term, appear to be higher in both Great Britain and Sweden than in the United States. (For the United States, the appropriate comparison is with regression 8. Measurement is not sufficiently precise to distinguish between autonomous wage pressure in Great Britain and Sweden. This weakens the hypothesis that the relatively unfavorable Phillips curve for the United States may be attributed to relatively great *autonomous* power of *U.S.* unions on wage changes. This, of course, is not to argue that the unions play no role in the problem, for their presence may be felt in the speed of wage adjustment, the relation between unemployment and vacancy rates, or the reaction of wages to price changes. Moreover, the constant terms in Table 1 will also reflect in part the constant term of the UV relation which can vary between countries and thus obscure differences in autonomous institutional wage pressures.

3) A second possible explanation for the relatively poor position of the *U.S.* Phillips curve is that the average rate of price inflation and/or the reaction of wages to price changes has been relatively high in the United States. During 1951–68 the average annual rate of change of prices was 4.3 percent in Sweden, 3.9 percent in Great Britain, and 2.3 percent in the United States. Thus, the hypothesis by Friedman noted in Section I does not explain the ranking of this set of Phillips curves. Indeed, the ranking of both the average rates of inflation *and* the sums of the distributed lag coefficients are exactly *inverse* to the average rate of price inflation over the period Nevertheless, the coefficients on the price change variables in this limited sample may provide some very tentative support for the accelerationist hypothesis. Under this hypothesis accelerating wage or price inflation is predicted only when aggregative policies are used to *maintain* an unemployment rate below the equilibrium rate for some period of time. Most industrialized nations have experienced sufficiently frequent cyclical fluctuations that they have not been able to generate data which would test this hypothesis persuasively. Sweden, however, provides a possible exception as the low and narrow range of unemployment rates indicates. In Table 1, the coefficient on the price change vari-

able is highest in Sweden, and when the distributed lag variable is used, the sum of coefficients was just above or just below unity in several instances (see regression 2 for example). Thus, institutional efforts or ability to protect real wages against the eroding influence of price increases are strongest in Sweden where the average rate of inflation has been highest among the countries in this sample (and where the Phillips tradeoff has been most favorable).

III. The Slope and Position of the UV Curve

It is clear from the empirical work in the previous section that neither autonomous wage changes nor expectational forces are responsible for the *relative* position of the *U.S.* Phillips curve. Thus the burden of explaining the international differences falls on the coefficient of the unemployment rate which is a hybrid of the wage flexibility coefficient and the forces influencing the position of the UV relation. Our attention will be limited to the hypothesis that the UV relation is relatively high in the United States. If comparable unemployment and vacancy data were available, a straightforward test of this proposition would be possible. However, job vacancy data are obscure or shaky for some countries, and regressions of the vacancy rate on the unemployment rate can at most show that the UV relation for the United States is displaced relatively far from the origin but will not yield information on specific characteristics of disequilibrium behavior in the labor market which are responsible for the finding. Since the possibilities are quite diversified and have different policy implications, the procedure in the subsequent empirical work will be to examine separately the factors which are shown to determine the UV relations.

1. *The Incidence of Unemployment*

The weekly flow of new unemployment is almost twice as high in the United States as in Great Britain or Sweden. Over the postwar periods for which data are available the average annual I_u are .63 for the United States, .33 for Great Britain, and .31 for Sweden. (The small size of these rates tends to obscure the massive flows they represent. At recent labor force levels the data imply that on average, four hundred seventy thousand people became unemployed each week in the United States. The comparable figures for Great Britain and Sweden are eighty-two thousand and twelve thousand.) This finding is a considerable surprise, both because it indicates that the incidence of unemployment may be responsible for most of the observed international differences in the unemployment rate[13] and because most nonmacroeconomic labor market policy during the past decade has emphasized the reduction of unemployment and unemployment differentials via programs which are largely designed to reduce the duration of unemployment.

Reference to equation (4) indicates that the relatively high I_u that characterizes labor market equilibrium in the United States must result from higher turnover flows, a lower level of labor market efficiency in the sense that a larger proportion of each turnover flow may incur unemployment in the United States,[14] or a labor

[13] Using 1965 data and the steady-state definition given by equation (2) we may divide the *U.S.* 1965 unemployment rate by the rate of incidence of unemployment for that year to obtain an estimate of the average duration, 7.3 weeks. This average duration may be multiplied by the observed 1965 rate of incidence in Great Britain and Sweden to obtain the aggregate unemployment rates estimated to prevail *if* the average amount of time between jobs was the same (7.3 weeks) in each country. For Great Britain the estimate is 2.2 percent and for Sweden the estimate is 1.8 percent—in each case quite close to the observed rate for the year.

[14] The use of the word "efficiency" is intended to include the possibility that the stock and quality of labor market information is sufficiently imperfect or the cost of search is sufficiently low that a larger fraction of gross turnover in the United States finds unemployed search behavior rational.

TABLE 2—WEEKLY FLOW OF NEW UNEMPLOYMENT AS PERCENT OF LABOR FORCE, UNITED STATES, 1970

Occupation	I_u	Industry	I_u
White Collar	.31	Agriculture	.96
Professional and Managerial	.16	Construction	1.11
Clerical	.48	Manufacturing	.58
Sales	.48	Durable	.56
Blue Collar	.67	Nondurable	.60
Craftsmen	.40	Transportation and Public Utilities	.34
Operatives	.79	Trade	.66
Laborers (nonfarm)	1.09		
Service	.63		
Males		Females	
Age 16–19	2.24	Age 16–19	2.56
20–24	1.06	20–24	1.11
25–44	.30	25–44	.61
45+	.20	45+	.31

Source: Data computed from Paul Flaim and Paul Schwab, Tables A-7, A-9, A-10, A-14, and A-15.

force structure in which sectors with high turnover in all economies are relatively important. In the remainder of this section, these alternative possibilities will be assessed.

It is clear from the data in Table 2 that the composition of the labor force has little influence on the relatively high flow of new unemployment in the United States. In virtually all sectors I_u substantially exceeds the aggregate value for Great Britain and Sweden, so that even if each country had the same industrial and occupational structure, I_u would be relatively high in the United States. As it turns out, the actual structural differences are minor. Moreover, these findings cannot be "explained" by the racial heterogeneity of the *U.S.* labor force. Although the weekly rate of new unemployment is much higher among nonwhites—for males, it was responsible for the full racial difference in unemployment rates in 1970—the I_u for whites (.42 for males and .70 for females in 1970) is still significantly above the average rates recorded abroad.[15]

[15] Data for the computations supporting these statements are from Flaim and Schwab.

On the other hand, the data in the bottom panels of Table 2 indicate that women and teenagers are the main source of the relatively high I_u in the United States. The rate of new unemployment for males with a primary attachment to the labor force is no higher than the average I_u observed abroad. Thus, if women and teenagers were a relatively large proportion of the *U.S.* labor force, the high amount of turnover-related unemployment could be explained. However, the demographic composition of the labor force in each of the countries under study is quite similar. During the late 1960's, youth (14–24 years) constituted about 20 percent and females 35–40 percent of the labor force in each country. (See *OECD*, 1968.) Moreover, the demographic differences in I_u reflect more than the racial composition of the various demographic groups. If one assumes (with strong support from the labor force survey data) that the average duration of unemployment is the same for both races, it is possible to compute the probable I_u for each of the age-sex cells in Table 2 for an all-white labor force. But with this adjustment for racial hetero-

geneity, the aggregate I_u falls by only .035 percent of the labor force. In the absence of evidence that labor force structure accounts for the large flow of new unemployment in the United States, turnover and labor market efficiency are examined below.

There is some direct evidence that the turnover flows are larger in the United States. During the postwar years for which data are available, the average monthly separation rate was 4.3 percent in the United States, 2.5 percent in Sweden, and 2.6 or 3.5 percent in Great Britain, depending on the series chosen.[16] Thus, even if the α, β, and γ coefficients are the same in each country, the incidence of unemployment would have been greater in the United States, neglecting the unlikely possibility that the gross labor force entry rate (which is not recorded in the two European countries) was sufficiently lower in the United States to offset the higher separations rate.

The reasons for the differences in turnover rates are not easy to isolate. Although there have been predominantly qualitative discussions of widespread labor hoarding and employer paternalism in many European countries and assertions that the United States has a "more mobile" society, these and a number of alternative economic explanations of higher American turnover rates cannot be tested carefully in the absence of quit and layoff data for European countries. On a more descriptive level, one would expect relatively rapid shifts in the structure of employment demand and relatively wide annual seasonal fluctuations to induce relatively high turnover rates. Structural shifts in demand will have a direct influence on layoffs as well as an indirect influence on the level of quits, for rapid structural change implies a relatively rapid obsolescence of job information and thus a greater expected return to repeated information search.

An index of structural change was computed using employment data for nineteen two-digit manufacturing sectors.[17] By this measure, the United States did experience more structural change than Great Britain and Sweden during the 1951–60 period. However, during the 1960's, the same measure indicates that Sweden experienced considerably more structural change than either the United States or Great Britain, although, as Table 2 indicates the *U.S.* I_u and separations rate were still relatively high during this period despite less structural change. When the annual variance of monthly seasonal adjustment factors for unemployment is used to measure seasonality over the period 1956–67, the average annual variance is 1780 for Sweden 103 for the United Kingdom, and 113 for the United States.[18] This evidence also is not

[16] For the United States and Sweden, the data cover the mining and manufacturing sector (see U.S. Bureau of Labor Statistics and Statistiska Centralbyrån). The lower British estimate is from the British Ministry of Labour series which excludes people who begin and end their employment within a given month, so that the series understates the separations rate as it would be measured in the United States and Sweden. The second British series, unpublished but compiled by the Inland Revenue Department, refers to the total number of separations within the field of pay-as-you-earn income tax divided by the number of jobs in that category. Although the coverage of this series is therefore broader than the American and Swedish data, the bias, if any, is unclear, for it includes high turnover sectors such as construction along with the civil service.

[17] The index of structural change used in the regressions as a measure of long-run shifts in demand is

$$B = \sum_{i=1}^{8} \mid \Delta E_{it} - \Delta E_t \mid \frac{E_{it}}{E_t} (100)$$

where E_t = total employment in year t
E_{it} = employment in industry i in year t
ΔE_{it} = the rate of change of employment in the ith industry in year t
ΔE_t = the rate of change of employment for all industries

The index of structural change is computed from annual data on total employment in eight one-digit industrial sectors obtained from *OECD*, 1964, 1968.

[18] The numbers were computed from unpublished data supplied by the U.S. Bureau of Labor Statistics and the *OECD* Statistical Office.

TABLE 3—RATE OF INCIDENCE (I_u) REGRESSIONS, UNITED STATES, GREAT BRITAIN, SWEDEN

Country and Time Period	Regression No.	Regression Coefficients: Constant	U	N_u	B	σ^2	R^2	d
United States (1957–66)	9	.57	.012				.19	1.45
		(.05)	(.009)					
	10	.72	.015		.0040	−.0015	.67	2.78
		(.07)	(.008)		(.0110)	(.0006)[a]		
	11	.73		.015	.0044	−.0015	.66	2.75
		(.07)		(.008)	(.0111)	(.0006)[a]		
Great Britain (1957–67)	12	−.05	.118				.40	1.79
		(.14)	(.045)[a]					
(1957–66)	13	.28	.078		−.046	−.0013	.47	1.45
		(.15)	(.039)[b]		(.039)	(.0011)		
	14	.33		.073	−.044	−.0014	.39	1.47
		(.16)		(.046)	(.043)	(.0012)		
Sweden (1962–68)	15	.17	.075				.50	2.28
		(.16)	(.0346)[b]					
	16	.20		.0764			.42	2.26
		(.06)		(.0404)[b]				

Sources: U.S. Department of Labor; Ministry of Labor *Gazette; SCB* 1969.
Note: B is an index of structural shifts in employment defined in the text; σ^2 is the variance in monthly seasonal adjustment factors; N_u is the number of unemployed. Standard errors are in parentheses and d is the Durbin-Watson statistic.
[a] Significant at 5 percent level.
[b] Significant at 10 percent level.

fully consistent with the ranking by labor turnover.

2. *The Cyclical Sensitivity of New Unemployment and New Vacancies*

In the absence of data on quit, layoff, and gross labor force entry rates for each country, the labor market efficiency parameters cannot be directly determined. However, some indication of their relative magnitude is possible from an analysis of the cyclical behavior of I_v and I_u, which itself is germane to the position of the UV and Phillips curves.

The cyclical sensitivity of I_u is described by the estimates in Table 3. The regressions in Table 3 are on generally comparable annual data, although the number of observations is small and the degrees of freedom problem is particularly severe for Sweden. Since the rate of unemployment, U, is used as a proxy for Q, direct estimates of the cyclical coefficients may be biased, because I_u is a behavioral function of the unemployment rate (used as a proxy for excess demand) as well as one of its determinants (see equation (2)). Therefore, reduced form estimates in which $I_u = f(N_u)$, where N_u is the number of people unemployed longer than one week as a percent of the labor force, are included with the direct estimates in Table 3. Finally, both linear and non-linear relations were tested, but the reported linear relations performed best.

Regressions 9, 12, and 15 provide comparable estimates when the effect of all other influences is assumed to be in the constant term.[19] In Great Britain and Sweden the incidence rate varies directly with the unemployment rate, indicating that changes in new unemployment are

[19] Similar estimates can be computed from the reduced form regressions 11, 14, and 16.

more closely associated with changes in layoffs than with changes in quits or labor force entry. In the United States, however, the rate of incidence is not significantly related to the rate of unemployment, indicating that in tight labor markets the decline in new unemployment associated with a reduction in layoffs is offset by a rise in the flow of new unemployment associated with the (increasing) voluntary flows of quits and labor force entrants.

It has been suggested that with strong demand pressure the unemployment induced by a rising quit rate could increase sufficiently to generate a positively sloped section in a Phillips curve (see fn. 2). However, for this result to obtain, 1) the percent of the labor force unemployed through "voluntary" choices, U_V, should be negatively related to the percent of the labor force unemployed because of layoff, U_L, and 2) the absolute value of the coefficients on U_L must be greater than unity. The evidence indicates that these conditions do not hold for the United States. A regression on quarterly data for the period 1967: II (when these series were first published) through 1970: IV yields

$$\begin{aligned} U_V = \underset{(.150)}{1.448} + \underset{(.085)}{.557}\, U_L - \underset{(.106)}{.373}\, \text{DUM1} \\ + \underset{(.095)}{.156}\, \text{DUM2} + \underset{(.096)}{.384}\, \text{DUM3} \\ R^2 = .87 \end{aligned}$$

when quarterly dummies are included. Clearly the hypothesis fails, for the sign indicates that the voluntary unemployment rate varies directly with the involuntary rate. Whatever the influence of voluntary turnover on the rate of new unemployment, these flows are more than offset by the falling average duration of unemployment in tight labor markets.[20]

The remaining regressions in Table 3 were run to isolate specific noncyclical determinants of the incidence of unemployment. These determinants should reflect variations in employment and labor force entry which are independent of movements in aggregate demand. If there are marked year to year changes in these indices which are largely independent of cyclical fluctuations, their influence should be reflected in a time-series regression. The indicators of noncyclical factors affecting the amount of turnover during the course of a year are an index of changes in the industrial composition of demand (see fn. 17) and the seasonal variance of unemployment over the year.

Regressions 10 (for the United States) and 13 (for Great Britain) are run on this data. In each equation, the role of the unemployment rate is unchanged, although it is now measured with less precision in the case of Great Britain, and the index of structural change is not significant. The coefficients on the index of seasonality are negative in both cases and measured with considerable precision in the case of the United States. This is the opposite of the expected effect and the reason for it is not immediately clear.

We are left with two salient differences between the United States and the sample European countries: the weekly flow of unemployment is relatively high and cyclically insensitive in the United States. From equation (4) it is clear that the level of I_u depends on the size of labor turnover flows and the proportion of each incurring some unemployment. Also, from equation

[20] For example, the quit rate, QR, is *negatively* related to unemployed quits as a percent of the labor force, U_Q.

$$\begin{aligned} U_Q = \underset{(.103)}{1.020} - \underset{(.050)}{.210}\, QR - \underset{(.037)}{.012}\, \text{DUM1} \\ - \underset{(.037)}{.012}\, \text{DUM2} + \underset{(.072)}{.318}\, \text{DUM3} \quad R^2 = .76 \end{aligned}$$

Data for this regression and the one in the text are from U.S. Bureau of Labor Statistics.

TABLE 4—SEPARATIONS RATE REGRESSIONS, UNITED STATES, GREAT BRITAIN, SWEDEN

Country and Time Period	Regression Number	Constant	Regression Coefficient U	R^2	d
United States (1951–66)	17	5.62 (.31)[a]	−.28 (.06)[a]	.57	1.10
Great Britain (1951–66)	18	3.11 (.17)[a]	−.19 (.06)[a]	.41	.86
Sweden (1958–68)	19	3.90 (.20)[a]	−.71 (.10)[a]	.85	1.30

Sources: U.S. Bureau of Labor Statistics, 1968; Ministry of Labour *Gazette;* Konjunkturinstitutet.

[a] Significant at the 1 percent level.

(6) it is clear that where

$$\frac{dI_u}{dQ} < 0 \text{ as in Great Britain and Sweden:}$$

$$|\alpha q'| + |\gamma p'| < |\beta y'|$$

$$\frac{dI_u}{dQ} = 0 \text{ as in the United States:}$$

$$|\alpha q'| + |\gamma p'| = |\beta y'|$$

Therefore, if the labor market efficiency parameters are the same in all countries, i.e., $\alpha=\beta=\gamma$, the involuntary elasticity must exceed the voluntary elasticity in European countries.

However, this inference is contradicted by Table 4, which reports the results of regressions of the separations rate on the unemployment rate. In each country, the coefficient on the unemployment rate is negative and measured with high precision, indicating that quits are more sensitive to cyclical variations than layoffs. It is clear, therefore, that the labor market efficiency parameters are not equal. It is also clear that the relatively high *U.S.* separations rate cannot be explained by a relatively weak pressure of demand. Reductions in the unemployment rate to common European levels would further increase the separations rate. The inverse relation between the separations rate (vacancy incidence rate) and the unemployment rate is strongest in Sweden, indicating that the cyclical sensitivity of quits relative to layoffs is greatest there.

If either the quit or layoff elasticity were known for each country, the other could be determined from the coefficients in Table 4. As noted above, there is considerable qualitative evidence indicating that layoff elasticities are relatively high in the United States. One indication of differences in the cyclical variation in layoffs among countries is the speed with which employment adjusts to changes in output, and employment adjustment coefficients estimated by Frank Brechling and Peter O'Brien from postwar data for several countries support this view. The values of the employment adjustment coefficient (.590 for the United States, .219 for the United Kingdom, and .285 for Sweden) indicate that y' is two to three times greater in the United States. Thus the layoff elasticity in the United States is two to three times that observed in Europe.

If one assumes that the employment adjustment coefficients represent the *relative* size of y' in each country, the actual value for each country can be determined if the cyclical layoff elasticity can be deter-

mined in any one of the countries. Direct estimation is possible in the United States, and regression of the layoff rate on the unemployment rate for the period 1951–68 yields an estimate of $y'=.29$, which implies a y' of .11 in the United Kingdom and .13 in Sweden.

Once y' is determined for each country, q' can be inferred from the fact that the regression coefficients on the unemployment rate in Table 4 should approximate $(q'+y')$. Applying this formula yields estimates of q' of $-.45$ for the United States, $-.39$ for Great Britain, and $-.70$ for Sweden. Since q' is found to be greater than y', the main difference in the cyclical sensitivity of I_u is to be found in the proportion of voluntary and involuntary turnover which incurs unemployment. For Sweden and Great Britain:

$$|\alpha q'| < |\beta y'| - |\gamma p'| \tag{9}$$

and for the United States:

$$|\alpha q'| = |\beta y'| - |\gamma p'| \tag{10}$$

Since we know that q' exceeds y' for each country, unemployment due to quits must be less than unemployment due to layoffs, so $\alpha<\beta$ is required.

We have said very little about labor force entry and reentry as yet. Unfortunately, data on gross labor force flows, the variable of main relevance here, are scarce both in the United States and abroad. If one assumes that gross and net labor force flows are correlated, and there is some fragmentary evidence in support of this proposition for the United States,[21] then p' should vary directly with the size of the labor force participation elasticity. Evidence for the three countries under study indicates that the labor force elasticity is highest in Sweden and lowest in the United States. (See Flanagan.) In terms of equations (9) and (10), this result reinforces the conclusion that a higher fraction of voluntary turnover flows in the United States than in Europe must incur unemployment.

To summarize the findings of the past two sections, the level of the incidence of unemployment is higher in the United States in part because the underlying turnover flows are higher irrespective of cyclical conditions and in part because the proportion of voluntary turnover flows incurring unemployment appears to be higher in the United States. It is not clear with the available data whether the high turnover rate is due to a relatively high quit rate, layoff rate, or both. There is qualitative evidence indicating that at least the layoff rate is higher in the United States. Information on the quit rate is less reliable, and the particular labor market inefficiencies responsible for the relatively high proportion of voluntary turnover incurring unemployment are also unclear.

3. *The Duration of Unemployment*

The evidence in the preceding sections indicates that international variations in the flows of new unemployment and new vacancies alone imply a relatively high UV relation and thus Phillips curve for the United States. It remains to discover whether the influence of the flow of new unemployment is reinforced by the average time spent between jobs in the sample countries. (Data limitations preclude study of the average duration of vacancies.) Here the distribution of unemployment by duration is important, for each of the countries complains of structural problems, and at times the relatively high unemployment rates in the United States

[21] Data on gross labor force flows in the United States are published for the early postwar period only. For 1949–52, we find

	1949	1950	1951	1952
Additions to labor force (LF) as a percent of LF	4.5	4.7	5.2	5.1
Unemployment rate:	5.9	5.3	3.6	3.1

Source: U.S. Bureau of the Census.

have led casual observers to infer that these problems are relatively severe in the United States. If this view were correct, one would expect to find a relatively large and cyclically insensitive amount of long duration unemployment in the United States, implying that bottlenecks will be reached relatively early in an expansion and that the gap between actual and desired (equilibrium) employment will be relatively large. This view can be tested by observing U_d, the rate of long duration unemployment (i.e., the number of people unemployed thirteen weeks or more as a proportion of the labor force, except in the United States where published data are for fifteen weeks and over), and by estimating its cyclical sensitivity.

During the postwar period, long duration unemployment as a proportion of total unemployment has varied from 12 to 32 percent in the United States, 25 to 40 percent in Great Britain, and 11 to 22 percent in Sweden. Over the same period U_d, the rate of long duration unemployment has ranged between .33 and 2.14 percent in the United States, .48 and 1.43 in Great Britain, and from .30 to .62 in Sweden.[22] The width of the range is dictated by the amplitude of cyclical fluctuations and the cyclical sensitivity of long-term unemployment, but the minimum levels of the long duration unemployment rate do not indicate that the United States has a relatively difficult structural problem. But given the different flows of new unemployment, the lowest American long duration unemployment rates—.33 in 1953, .37 in 1952, and .48 in 1951—occurred in years when the aggregate unemployment rates were twice as large as rates commonly experienced in Sweden and Great Britain.

The cyclical behavior may be studied by estimating the coefficients of the following equation:

$$(11)\quad U_d = a_0 + a_1 U_t + a_2 U_{t-1} + a_3 t + e$$

in which U_d is the rate of long duration unemployment, U is the aggregate unemployment rate, t is the time trend, and e is the error term. The cyclical sensitivity of the long duration unemployment rate will be reflected in a_1 and a_2, while a_3 will indicate whether there has been secular change in the long duration rate, apart from cyclical influences. Since the current rate of long duration unemployment is the result of past as well as current labor market conditions, the lagged unemployment rate can also be introduced. (Also, reduced form estimates can be computed from regressions of U_d on $(U-U_d)$ and t.)

The parameters of equation (11) were estimated from annual data for the United States and Great Britain and from quarterly data for Sweden (where the only comparable and consistent series provide only eight annual observations). For the United States and Great Britain, both the structural and reduced form equations were estimated, and the results are provided in Table 5. For Sweden, the use of quarterly data enabled a more precise isolation of the lagged response of long duration unemployment to changing labor market conditions. The results of several experiments indicated that a lag of one quarter was optimal. This equation, which includes the seasonal dummies $D1$, $D2$, and $D3$, appears as regression 28 in Table 5.

There are two findings of interest. The first is that the cyclical sensitivity in the long duration unemployment rate is *greatest* in the United States (see regressions 20, 24, and 28). This conclusion stands from both the direct and reduced form estimation of a_1. The estimate of a_1 for the United States computed from regression 22 is .45 (the direct estimate in regression 20 is .46), and the reduced form estimate for Great

[22] Data are from *Manpower Report of the President*, Ministry of Labour *Gazette*, and Statistiska Centralbyrån. All data adjusted to *U.S.* definitions by procedures described in Flanagan, Appendix A.

TABLE 5—RESULTS OF DURATION REGRESSIONS, UNITED STATES, GREAT BRITAIN, SWEDEN

Country and Time Period	Regression No.	Constant	a_1	a_2	a_3	$\frac{a_1}{1-a_1}$	$\frac{a_2}{1-a_1}$	$\frac{a_3}{1-a_1}$	R^2	d
United States (1950–68)	20	−1.08	.46		−.0008				.97	2.47
			(.02)[a]		(.0041)					
	21	−1.15	.45	.032	−.0013				.97	2.30
			(.03)[a]	(.022)	(.0040)					
	22	−1.86				.82		−.0009	.90	2.33
						(.06)[a]		(.0076)		
	23	−2.07				.77	.10	−.0008	.92	2.06
						(.07)[a]	(.07)	(.0073)		
Great Britain (1951–67)	24	−.37	.41		.013				.93	1.48
			(.046)[a]		(.005)[c]					
	25	−.67	.37	.16	.008				.98	1.92
			(.02)[a]	(.03)[a]	(.003)[b]					
	26	−.41				.56		.026	.81	1.42
						(.12)[a]		(.008)[a]		
	27	−.48				.52	.094	.023	.83	1.43
						(.12)[a]	(.077)	(.008)[b]		
Sweden (1962:I–68:IV)	28	.056	.22	.004	−.002	−.017	−.109		.72	1.66
			(.05)[a]	(.022)[d]	(.048)	(.049)	(.048)[c]			

Sources: See text and sources listed for Table 4.
[a] Significant at 1 percent level.
[b] Significant at 2 percent level.
[c] Significant at 5 percent level.
[d] Significant at 10 percent level.

Britain is .36 (compared with the direct estimate of .41).

The regressions imply that with an aggregate unemployment rate of 4 percent in each country, the corresponding rates of long duration unemployment would be .76 in the United States, .94 in Sweden, and 1.27 in Great Britain. Therefore, the evidence favors the conclusion that the relatively high extreme rates of long duration unemployment which have been observed in the postwar United States are indicative of loose labor markets rather than of relatively severe structural problems.

Secondly, while there is no discernible time trend in the United States, evidence of a secular increase in the rate of long-term unemployment appears in the British and Swedish regressions. The coefficients on the time trend in the latter countries are measured very precisely, and are relatively insensitive to the structural or reduced form specification.

To summarize, in each of the three sample countries there are structural problems. The literature of economic policy in each country is replete with references to "the regional problem," inflexible wage structures, and skill bottlenecks when unemployment is high. However, the analysis indicates that structural unemployment (as measured by the rate of long duration unemployment) is not *relatively* severe in the United States, so that the American UV relation is neither relatively steep nor displaced relatively far from the origin because of structural problems. These general conclusions are supported by the data

for sectors of the *U.S.* labor force which are often viewed as important sources of structural unemployment as defined above.[23] For example, the share of the aged, nonwhites and blue-collar workers in long duration unemployment declined between 1961 and 1969 as the aggregate unemployment rate fell from 6.7 to 3.5 percent, whereas the share of white-collar and service workers increased. The share of teenagers did increase over this period to 18 percent of the long duration unemployed. However, most teenager unemployment was of shorter duration—almost one-third of all unemployment of less than five weeks occurred among members of this age group. Since most of this short-term unemployment is experienced by labor force entrants and re-entrants this group may have an important impact on the cyclical stability of the I_u statistic, although it is not evident at this time that this unemployment reflects normal allocative frictions.[24]

IV. Conclusion

The results reported in this paper confirm that the *U.S.* Phillips curve presents a less favorable tradeoff for monetary and fiscal policies than wage inflation-unemployment relations observed abroad. Neither the autonomous influence of institutions on wages nor the influence of expected price changes are important explanations of the *relative* position of the *U.S.* curve during the postwar period. Instead, the *U.S.* tradeoff reflects the fact that the United States exhibits the most incomplete adjustment of supply and demand in the labor market. Moreover, labor market maladjustment apparently increased, inducing an even poorer tradeoff, during 1967–70. However, the relatively poor adjustment is not due to relatively severe immobilities in the United States, but reflects a relatively high rate of separations and labor force entry along with the associated unemployment. The reasons for the relatively high gross turnover flows in the United States are as yet not well understood, and the traditional presumption that turnover is a by-product of allocational efficiency may be particularly qualified by more thorough analysis of racial differences in gross flows.

These findings have some implications for labor market policies, which have generally been designed to narrow the gap between the vacancy and unemployment rates by reducing the average duration of unemployment or vacancies. To the extent that these programs are successful, they will shift a properly measured Phillips curve. However, in pursuing these programs, it should be recognized that they do not generally treat the main source of observed international differences in Phillips curves—the factors responsible for the high incidence of unemployment and vacancies in the United States at every rate of excess demand—except to the extent that they reduce the percentage of quits, layoffs, and labor force entrants which incur unemployment.

[23] Data in this paragraph are from *Manpower Report of the President*, pp. 225–28.

[24] For other evidence on this observation, see Hall.

REFERENCES

R. Bodkin, E. Bond, G. Reuber, and T. R. Robinson, *Price Stability and High Employment*, Ottawa 1967.

F. Brechling and P. O'Brien, "Short Run Employment Functions in Manufacturing: An International Comparison," *Rev. Econ. Statist.*, Aug. 1967, *49*, 277–87.

G. C. Chow, "Tests of Equality Between Sets of Coefficients in Two Linear Regressions," *Econometrica*, July 1960, *28*, 591–605.

B. Corry and D. Laidler, "The Phillips Relation: A Theoretical Explanation," *Economica*, May 1967, *34*, 189–97.

J. C. R. Dow and L. A. Dicks-Mireaux, "The Excess Demand for Labour: A Study of Conditions in Great Britain, 1946–1956," *Oxford Econ. Pap.*, Feb. 1958, *10*, 1–33.

R. J. Flanagan, "A Study of International Differences in Phillips Curves," unpublished Ph.D. dissertation, Univ. California, Berkeley 1970.

P. O. Flaim and P. M. Schwab, "Employment and Unemployment in 1970," *Special Labor Force Report* No. 129, U.S. Bureau of Labor Statistics, Washington 1971.

M. Friedman, "The Role of Monetary Policy," *Amer. Econ. Rev.*, Mar. 1968, *58*, 1–17.

R. E. Hall, "Why Is the Unemployment Rate So High at Full Employment?" *Brookings Papers on Economic Activity*, *3*, 1970.

B. Hansen, "Excess Demand, Unemployment, Vacancies and Wages," *Quart. J. Econ.*, Feb. 1970, *84*, 1–23.

R. G. Lipsey, "The Relation Between Unemployment and the Rate of Change of Money Wage Rates in the United Kingdom, 1862–1957: A Further Analysis," *Economica*, Feb. 1960, *27*, 1–31.

J. Mincer, "Comment," in *Measurement and Interpretation of Job Vacancies*, Nat. Bur. Econ. Res. Conference, New York 1966.

G. L. Perry, "Changing Labor Markets and Inflation," in *Brookings Papers on Economic Activity*, *3*, 1970.

E. S. Phelps, "Money Wage Dynamics and Labor Market Equilibrium," in Phelps et al, 124–66.

——— et al, *Microeconomic Foundations of Employment and Inflation Theory*, New York 1970.

G. J. Stigler, "Information in the Labor Market," *J. Polit. Econ.*, Oct. 1962, *70*, 94–105.

Statistika Centralbyrån (*SCB*), "Arbetskraftsundersökningarna, 1968," *Statistiska Meddelanden*, AM 1969:10, Stockholm 1969.

———, *Allmän Månadsstatistik*, various issues, 1954–69.

———, "Konsumentprisindextel, 1914–1967," *Statistiska Meddelanden*, H 1968: 1C, 80, 85; Series Stockholm 1968.

Organization for Economic Cooperation and Development, (*OECD*), *Labour Force Statistics*, 1956–66, Paris 1968.

———, *Manpower Statistics*, 1950–62, Paris 1964.

President's Committee to Appraise Employment and Unemployment Statistics, *Measuring Employment and Unemployment*, Washington 1962.

U.K. Ministry of Labour *Gazette*, various issues, 1952–69.

U.S. Bureau of the Census, *Current Population Reports, Series* P-50, Nos. 19, 31, 40, 45.

U.S. Bureau of Labor Statistics, *Employment and Earnings Statistics for the United States*, Bull. 1312-6, Washington 1968.

U.S. Department of Labor, *Statistics on Manpower*, Mar. 1969.

U.S. President, *Manpower Report of the President*, Apr. 1971.

Required Disclosure and the Stock Market: An Evaluation of the Securities Exchange Act of 1934

By GEORGE J. BENSTON*

The Securities Exchange Act of 1934 was one of the earliest and, some believe, one of the most successful laws enacted by the New Deal. The stock market crash in 1929 and the Great Depression provided the impetus for reform of the stock markets in the belief that weaknesses of the institutions and ineptitude and/or chicanery among brokers and bankers were partially responsible for the losses incurred by stockholders. Although many critics, reformers and congressmen wanted Congress to enact "blue skies" legislation that would require all securities sold and traded to be approved by the federal government, President Franklin Roosevelt preferred the concept of "disclosure" (see Francis Wheat (1967)). Rather than having the government approve or disapprove securities, corporations whose securities are publicly sold and traded are required to disclose a large amount of predominantly financial information to the Securities and Exchange Commission (*SEC*) who make these data available to the public. Indeed, the Securities Exchange Act is described in its title and usually referred to as a "disclosure statute."

Although the financial community generally opposed this legislation and the preceding Securities Act of 1933,[1] most brokers, investors and government officials probably would find it difficult to conceive of the successful operation of the stock markets without the Securities Acts. Yet the economic rationale for the regulation of the securities markets was not examined carefully before the legislation was passed (which is not surprising, given turbulent times) nor has it been since,[2] even though the Securities Act of 1934 was extended in 1964 to include most corporations whose stock is publicly owned. Such an examination of one important part of the law—the financial disclosure requirements—is presented here. This analysis is particularly timely because the *SEC* appears to be shifting its emphasis towards increasing the disclosure requirements of almost all corporations whose stock is traded in the markets.[3]

* Professor, Graduate School of Management and Center for Research in Government Policy and Business, University of Rochester. Thanks are due to Robert Hagerman and Stanley Engerman for their criticisms, some of which I agree with.

[1] A recent history by Ralph DeBedts reviews these events uncritically.

[2] George Stigler (1964a, b) and Irwin Friend and Edward Herman provide the first quantitative analysis of the Securities Act of 1933 of which I am aware. The only other analyses are in two papers which I published in 1969.

[3] A rather detailed "Report and Recommendations to the Securities and Exchange Commission" from the Disclosure Policy Study headed by Wheat recommends:

> . . . that for the future, greater attention be paid to those continuing disclosures which benefit the trading markets in securities. Prior to 1964 the continuing disclosure reached only those issues whose securities were listed on exchanges and those which had voluntarily registered securities under the '33 Act. Full exercise of that authority might have deterred listing. This is no longer the case, and a serious impediment to progress in disclosure policy has been removed. [Wheat Report, p. 11]

I. The Disclosure Requirements of the Securities Exchange Act of 1934

The '34 Act requires that a corporation whose stock is traded on a registered stock exchange or who registered a stock issue:

a) file detailed balance sheets, income statements, and supporting substatements (form 10K) within 120 days after the close of its fiscal year;

b) file a much less detailed semiannual report (form 9K) within 45 days after the first half of the fiscal year;

c) file a "current report" (form 8K) 10 days after the end of any month in which certain "significant" events occurred (such as a change of control of the corporation, material legal proceedings undertaken, material change of securities outstanding, and revaluation of assets).

In 1964, the disclosure requirements were extended to almost all corporations with at least 500 stockholders or $1 million in assets. (Exceptions are regulated companies whose statements were prescribed, such as banks and insurance companies). Thus, all but the smallest corporations now are covered by the Act.

Section 13(b) of the '34 Act (and section 19(a) of the '33 Act) gives the *SEC* the power to prescribe the form and content of the financial statements filed under the Act. In general, the *SEC* has followed generally accepted accounting procedures, although it has influenced these procedures by insisting that assets not be revalued upward, goodwill be amortized rapidly, and other "conservative" biases be reinforced. In this regard, the *SEC* has not followed the "disclosure rather than approval" philosophy of the Securities Exchange Act. This policy was established by Accounting Series Release No. 4 in 1938, which states that

> . . . where financial statements filed . . . are prepared in accordance with accounting principles for which there is no substantial authoritative support, such financial statements will be presumed to be misleading or inaccurate *despite disclosures* contained in the certificate of the accountant or in footnotes to the statements provided the matters involved are material. (italics added.)

Whether disclosure, as defined and required by the *SEC*, has been meaningful and beneficial, is the question asked here—not whether disclosure, as such, is good or bad.[4]

II. The Rationale Underlying the Legislation

It would seem that any argument against disclosure is equivalent to an argument for secrecy. But such is not the case. Prior to the passage of the Securities Exchange Act, corporations could disclose what they wished to their current and potential stockholders and, if they were listed on the New York Stock Exchange (*NYSE*), American, Chicago (Midwest), or other regional exchanges, had to submit balance sheets and income statements to the Exchange. For the year ended December 31, 1933, all *NYSE* corporations were audited by CPA firms, all listed current assets and liabilities in their balance sheets, 62 percent gave their sales, 54 percent the cost of goods sold, and 93 percent disclosed the amount of depreciation expense. These percentages had been increasing fairly steadily prior to 1933, although there was little change after 1928. (See Benston (1969b), p. 519.) One could argue (as did the *NYSE*), that the legislation was not needed.

One could also argue that the disclosure policy followed by corporations in the absence of legislation is in the best interests

[4] The possible detremental effect of the *SEC*'s definition of disclosure on the development of improved and innovative accounting procedure has been argued elsewhere (Benston, 1969b). While the issue is important to the question of the efficient operation of securities market, it is not emphasized in this paper.

of their stockholders. If management believed that the marginal revenue to the stockholders as a group from disclosure would exceed the marginal cost of preparing and supplying the information, they would disclose their financial and other data. The marginal revenue might include the savings to stockholders of not having to gather the data privately, the reduced cost of capital to the firm if prospective stockholders' uncertainty about the firm were reduced, improvement in the marketability of the firms' shares if investors desired financial information, etc. The marginal costs of disclosure might include the cost of preparing and distributing the statements, the costs incurred in informing competitors, suppliers, customers, and government officials, and the cost of misinforming stockholders when accounting statements report economic events incorrectly or inadequately (as when all research and development and advertising expenditures are charged to expense currently).

However, management might not issue financial statements (or might issue incomplete statements) if they underestimate the value of these statements to their current or potential investors, mismanage the corporation, intend to defraud investors, or if there are positive externalities in the efficient allocation of resources when all (or most) companies disclose financial data. Thus, one cannot immediately dismiss the argument that there is need for required disclosure solely by reference to the invisible hand of the market. Rather the question must be examined with respect to the rationale upon which the Securities Exchange Act of 1934 is based.

Underlying the disclosure requirements of the '34 Act is the belief that required disclosure of financial data is necessary for the fair and efficient operation of capital markets. The *SEC*'s 1969 Wheat Report (and most other writings on the subject) view disclosure as necessary to 1) prevent financial manipulation and 2) provide investors and speculators with enough information to enable them to arrive at their own rational decisions. (Wheat Report, p. 10.)[5] Perhaps even more important is the concept of "fairness," the belief that all investors, big and small, insiders and outsiders, should have equal access to relevant information. Whether these objectives can be achieved, a priori, by disclosure of financial data, and if they can, whether or not the evidence supports or rejects the hypothesis that they were, is considered in the balance of the paper. To facilitate the presentation of the material, fraud and manipulation is discussed first in Section III, followed by an empirical analysis of information and investors' decisions in Section IV, and the effects of the '34 Act on traded securities in Section V, on losses by stockholders in Section VI, and on investors' confidence in the market in Section VII.

III. Fraud and Manipulation

Fraud and manipulation may be of two different types with respect to disclosure. Published statements may contain false or misleading data or desired data may not be published at all but may be released in the form of news stories, rumors, etc., to manipulate the public's expectations and so affect stock prices. These are discussed in turn.

It is very difficult to determine whether the '34 Act prevented the publication of

[5] Also mentioned is ". . . the belief that appropriate publicity tends to deter questionable practices and to elevate standards of business conduct" (p. 10). These goals are inherently nonoperational, except as they refer to the prevention of fraud and manipulation, which is discussed below, or perhaps to insider stock dealings, which is not affected by the disclosure of financial data and hence is outside the scope of this paper. However, see Henry Manne for one view contrary to the *SEC*'s position.

fraudulent or misleading financial statements or even whether much fraud existed to any greater extent before or after the passage of the Act. In a situation of personal fraud by self-dealing or simple defalcation, required disclosure is of little value. Certified Public Accountants insist they they do not audit explicitly for fraud nor does the *SEC* ask them to do this, although I believe a good case could be made for this requirement. With respect to fraudulently prepared financial statements, I have reviewed such evidence as exists in another article (Benston (1969a), pp. 51–55). A search of the available literature, including the Senate and House hearings on the proposed securities legislation, fails to reveal much evidence of fraud in the preparation or dissemination of financial statements prior to 1934. For example, Wiley Rich, the author of a comprehensive survey of the legal responsibilities of accountants, states that "An extensive search has revealed not a single American case in which a public accountant has been held liable in a crime for fraud" (p. 100). It appears that the feeling in the early 1930's that published accounting statements were fraudulent or misleading was based on the "exposés" of William Ripley and others of the behavior of some large corporations at the turn of the century. Further, the recent BarChris, Yale Transport, Green Department Store, Continental Vending, and other cases show that fraud in financial statements, while relatively rare, has not been banished from the land.

The lack of evidence on fraudulent financial statements does not imply that published financial statements were or were not misleading. Prior to the passage of the Securities Act, it was very difficult for third parties, such as prospective stockholders, to sue accounting firms for negligently prepared financial statements. The courts held, under the rule of privity, that these reports were prepared for management only. (See Landell v. Lybrand.)

However, accountants were (and still are) liable for fraud " . . . if their audit has been so negligent as to justify a finding that they had no genuine belief in its adequacy for this again is fraud." (See Ultramares Corp. v. Touche, Niven and Co. N.Y. p. 185, N.E. p. 488.) The Securities Act changed accountants' liability dramatically, and now an investor may sue an accountant if, having relied on false or misleading statements, he "shall have purchased or sold a security at a price which was affected by such statement" (Section 18). It is important to note that the accountant must prove that the investor's loss was not a consequence of the financial statements rather than that the investor prove that he actually was misled by or even saw the statements.

In contrast to the lawmakers' expectations, an important consequence of this change in the law and of the *SEC*'s administration of the Acts appears to be that financial statements are more misleading than they were. The considerable liability of accountants under Section 18 has contributed to accountants following conservative, often worthless practice, since it is difficult to sue them successfully for preparing misleading statements if they follow traditional procedures. In addition, the *SEC* has insisted on historically based accounting, discouraging price level and other revaluation of assets and liabilities, refusing to permit publication of sales and income projections and other valuable economic data, etc. Thus published financial statements are more misleading than they otherwise might have been. Although accountants might not have made much progress in reporting the economic position and progress of corporations had there been no Securities Act, there is no empirical or a priori basis for an assertion that the '34 Act has had a net positive

effect on the publication of fraudulent or misleading financial statements.

While there is little direct evidence that corporate managers issued fraudulent financial statements, they may have refused to disclose information to create a climate in which they could manipulate stock prices by means of "pools" and "bear raids." At least such is the opinion of the *SEC*. In their booklet, *A 25 Year Summary of the Activities of the Securities and Exchange Commission, 1934–1959*, they say these practices

> . . . resulted in a situation in which no one could be sure that market prices for securities bore any reasonable relation to intrinsic values or reflected the impersonal forces of supply and demand. In fact, the investigation record demonstrated that during 1929 the prices of over 100 stocks on the New York Stock Exchange were subject to manipulation by massive pool operations. One of the principal contributing factors to the success of the manipulator was the inability of investors and their advisers to obtain reliable financial and other information upon which to evaluate securities. . . . [pp. xv–xvi]

To test this assertion, the financial statements (as published in *Moody's Investors Services*) of those over 100 corporations whose securities were subject to pools, as revealed by the Senate Committee on Banking and Currency, were examined for years before, during, and after the pools. Table 1 gives the percentage that disclosed such important information as sales and cost of goods sold. (All except 4 corporations included in the 1929 pools and 2 included in the 1930 pools had otherwise complete balance sheets and income statements published in *Moody's*.) The percentages with respect to sales were almost the same as those found for all companies listed on the New York Stock Exchange and a little lower with respect to cost of goods sold. Thus, while "pools" and "bear raids" may or may not have been unfair to investors, it is clear that their operations (successful or not) owed little to the nondisclosure of accounting data.

IV. Information and Rational Decisions of Investors and Speculators

The second rationale for the disclosure requirements is to allow "investors [to] make a realistic appraisal of the merits of securities and thus exercise an informed judgment in determining whether to pur-

Table 1—Disclosure of Financial Data by Corporations Whose Securities were Subject to Pools

		Percentage Disclosing[a]	
Pool Year	Statement Year	Sales	Cost of Goods Sold
1929 Pools	1927	58	47
(103 corporations)[b]	1928	58	48
	1929	61	50
	1930	60	52
1930 Pools	1928	70	50
(30 companies)[c]	1929	70	53
	1930	70	60
	1931	63	70
1931 Pools	1929	67	50
(6 companies)	1930	67	50
	1931	67	50
	1932	33	50
1932 Pools	1930	100	100
(2 companies)	1931	100	100
	1932	50	100
	1933	50	100
1933 Pools	1931	67	42
(12 companies)[d]	1932	67	42
	1933	67	42

Source: Pools: U.S. Senate Committee on Banking and Currency, part 17, pp. 7948–50, Financial Data: *Moody's Investors Service*, various years.

[a] Corporations whose statements were not in *Moody's* (4 of the 1929 pools and 2 of the 1930 pools) are included as "nondisclosure."

[b] Number of securities = 105; 2 securities of two corporations were listed

[c] Number of securities = 31; 2 securities of one corporation were listed

[d] Number of securities = 13; 2 securities of one corporation were listed

chase them" (Securities Exchange Commission, 1967, p. 1). This rationale is based on a belief that the data required by the *SEC* are "information." That is, the financial statements must provide data about a corporation that affect investors' expectations about its future prospects and relative riskiness and that were not previously known, such that the information was completely discounted and impounded in the market price of the securities before the time of disclosure.

There is serious question whether the financial data approved by the *SEC* can provide the investor with information. The *SEC* does not allow current market valuation of assets, estimates of future sales or projection of the effects of discoveries, favorable regulatory rulings, public acceptance of new products and other economic events. Additionally, the present value of many important occurrences such as management changes, styling, advertising campaigns and other marketing strategies, changes in the competitive environment, and the like cannot be estimated very well even if the *SEC* allowed accountants or others to publish their efforts. While this information need not come to the public through the financial statements, the *SEC*s requirement that these statements be prepared in a specified form implies that they include some relevant quantitative financial information.

But even if financial statements do contain information that is of value to investors, the data may not be made available to the public before insiders (including bookkeepers, secretaries, accountants, typists, and printers) see and take advantage of them. In requiring the filing of financial statements, the *SEC* is caught in a choice between speed and accuracy (in the sense of reporting to the letter of the formal and informal regulations), a choice which is resolved in favor of accuracy. As is noted above, the annual reports (10K) need not be filed with the *SEC* until 120 days after the close of a corporation's fiscal year. Whether the statements that have been filed are meaningful to investors and sufficiently timely to be of value is, of course, an empirical question, to which I now turn.

A. *The Information Content of Published Financial Data—Financial Statements and Stock Prices*

If the *SEC*'s disclosure requirements are meaningful, the statements they require should contain information, and thus investors' expectations about a corporation's earnings and prospects, riskiness, relationship to other firms, etc., should be affected by the information. Since numerous studies show that the market adjusts rapidly to new information,[6] the effect, if any, of previously unexpected data published in the financial reports of a corporation (j) should be reflected in changes in its stock prices (ΔP_{jt}) in the period t when these unexpected financial data (F^*_{jt}) become publicly available. Other factors also may occur during the same period and must be accounted for. Principal among these are changes in general market conditions (ΔM_t), changes in expected dividend payments (which often are announced at the same time that earnings data are announced) (ΔD_{jt}), and changes specific to the corporation's industry (ΔI_{kt}). In summary,

$$(1) \quad \Delta P_{jt} = f(F^*_{jt}, \Delta M_t, \Delta D_{jt}, \Delta I_{kt}, U_{jt})$$

where U_{jt} are other unspecified factors. This model was specified and tested, and the results were reported in a paper I published in 1967. They are summarized very briefly here. The specified model whose description follows was used for these tests, used in another study reviewed be-

[6] See Eugene Fama for a summary and review.

low (Ray Ball and Phillip Brown), and in further tests described below in Section V.

A two-stage estimating procedure was used that is based on a "market model" originally suggested by Markowitz and first applied in a context similar to the present study by Eugene Fama et al.

$$\tilde{r}_{jt} = \alpha_j + \beta_j r_{Mt} + \tilde{u}_{jt} \tag{2}$$

where $\tilde{r}_{jt}$ is the rate of return for security j for month t (defined as $ln((P_{jt}+D_{jt})/P_{jt-1})$ where P_{jt} is adjusted for stock dividends and splits and D_{jt} is dividend payments declared); r_{Mt} is a similarly measured rate of return on a market index M; α_j and β_j are parameters that can vary from security to security; and $\tilde{u}_{jt}$ is a random disturbance.[7] Tests applied by Fama et al., and Marshall Blume (1968, 1970) indicate that equation (2) ". . . is well specified as a linear regression model in that (i) the estimated parameters α_j and β_j remain fairly constant over long periods of time [e.g., the entire post-World War II period in the case of Blume], (ii) r_{Mt} and the estimated $\hat{u}_{jt}$ are close to serially independent, and (iii) the $\hat{u}_{jt}$ seem to be independent of r_{Mt}" (Fama, p. 403). The $E(\hat{u}_{jt}|F^*_{jt})$ may not$=0$. This is what was tested.

Equation (2) was applied in my 1967 paper to a sample of 483 companies traded on the New York Stock Exchange in 1964 (the latest year data were available at the time of the study). These included almost all *NYSE* traded companies from whom annual and quarterly financial data were available on the Compustat tapes. The statistics $\hat{\alpha}_j$ and $\hat{\beta}_j$ were computed from data that excluded months when the financial statements were made public. These statistics were then "plugged" into equation (2a) to compute the residuals, $\hat{u}_{jt}$, for the months when the financial data were published:

$$\tilde{r}_{jt} - \hat{\alpha}_j - \hat{\beta}_j r_{Mt} = \hat{u}_{jt} \tag{2a}$$

The $\hat{u}_{jt}$ computed with equation (2a) were regressed on estimated unexpected financial data, unexpected dividends, and variables that accounted for industry effects.

An expectation model had to be specified in order to estimate the unexpected financial data (F^*_{jt}). Since the *SEC*'s regulations require disclosure of accounting statements for prior periods as well as the present period,[8] expected financial data were taken to be a function of previously published data. Three averages of past data (a weighted average with geometrically declining weights determined by a variant of the Koyck transformation, and simple averages of three and five years' data) and a naive model (previous year's data) were used. Unexpected financial data F^*_{jt}, then, is measured by the difference between the reported and the expected data. Although the *SEC* did not require companies to report their quarterly data until recently, such data were required by the *NYSE*. Since investors' expectations would be affected by these reports, their effect is estimated by including third quarter financial data in the regressions in forms analogous to those used for the annual data.[9]

Since it is not clear which specific financial data investors use, the financial variables were defined alternatively as sales, cash flow net income, current operating income, and net income after extraor-

[7] Fama shows that this model is consistent with the Sharpe-Markowitz expected return model.

[8] The *SEC* requires that companies who float a stock issue under the provisions of the Securities Act of 1933 publish at least five years of comparative data, three of which must be certified by an independent certified public accountant.

[9] There is some possibility that these third quarter data could be collinear with the annual data, which might reduce the significance of the measured coefficients. I am indebted to Ross Watts for pointing out this possible error.

dinary gains and losses. Thus sixteen regressions were computed for the month when the financial data were sent to the *SEC* and stockholders (four expectations models and four definitions of income).

The regressions reveal that, except for the sales definition of financial data, none of the financial data variables in any of the expectations forms has a greater than minimal economic relationship to changes in stock prices, although the coefficients estimated are statistically significant. On the average, a 100 percent unexpected increase (or decrease) in the rate of change of income is associated with a 2 percent increase (or decrease) in the rate of change of stock prices in the month of announcement. Similar regressions were computed for the month when earnings were announced (which usually is a month before the *SEC* receives them), and similar results were found. The findings are not very dependent on the form of the expectations model used, although the naive model, where last year's rate of change is expected this year (a sort of "no-expectations" model), performed the best. This finding is contrary to the *SEC*'s requirement that companies provide comparative data for several past years. Thus, I conclude that this evidence is not consistent with the underlying assumption of the legislation, that the financial data made public are timely or relevant, on average.

Corroborative findings are reported by Ball and Brown, whose study used a variant of the model described above in equation (2). Their sample consists of 261 firms who reported on an annual basis. For most of these firms, data were available in the period 1946 through 1968 on the Compustat tapes. Ball and Brown considered each firm's year as an independent observation. They separated the data into two samples, one consisting of years in which income increased and the other of years in which income decreased unexpectedly, as determined from an expectation model in which a firm's expected income is a function of the income reported by the other firms. However, they did not account for dividend changes or for the information content of quarterly reports. Ball and Brown found that cumulatively, over a year, firms that show greater (or lesser) than expected income changes averaged greater (or lesser) than normal stock price returns. However, they calculate that ". . . of the value of information contained in reported income no more than 10 to 15 percent . . . has not been anticipated by the month of the [preliminary] report . . ." and compute that the marginal monthly rate of return of buying or selling short a portfolio of stock based on the information contained in the preliminary report is less than 1 percent (pp. 175–76). By the time of the final (*SEC*-required) report, there is almost no information that has not already been impounded in the price of the stock.

Another study of interest was made by William Beaver, who measured the volume of trading in the weeks before and after the announcement of preliminary earnings. He analyzed weekly price and volume for a sample of 143 firms (those included on the Compustat tape, traded on the *NYSE*, with other than December 31 fiscal years, with no dividends announced in the same week as the announcement, no stock splits in the quarters around that date, and with relatively few other "announcements"). A model similar to that given in equation (2) was used to abstract from market effects. Beaver found that the variance, average volume, and residual positive or negative return are much greater in the week that the preliminary reports are announced (without regard to the sign or magnitude of the earnings) than in the previous or following eight weeks. Further, the greater than normal activity appears to continue for about two to

three weeks after the announcement date, but with a considerably smaller magnitude. Contrary to the other studies, Beaver's results seem to indicate that financial statements do contain information or are at least used in some way. One interpretation is that investors use the announcements of preliminary earnings to switch their portfolios. Another is that brokers use the earnings reports as an excuse for "churning." A third explanation, one consistent with the findings of Robert Hagerman on the effect of requiring commercial banks to report earnings to the public in forms specified by regulatory agencies, is that while financial statements, as such, provide information, required reporting adds nothing. (Recall that all *NYSE* companies published some form of statement before passage of the '34 Act.)

One other study has been published that considers directly the usefulness of published annual financial statements in purchasing securities. Richard McEnally tested the value of using low price/earnings (P/E) ratios to choose portfolios of stocks. He used a randomly selected sample of 100 calendar-year stocks from which five portfolios of 20 stocks each were chosen for each of five years, 1961–65. High or low P/E ratios were the basis for forming portfolios. The stocks are assumed to be purchased on April 1 and held for a year. When the earnings used in the P/E ratio were those of the previous year (E_{t-1}), the portfolio with the lowest P/E_{t-1} outperformed the others slightly in terms of mean return, there being little difference among the other four portfolios.[10] McEnally then used the ratio of price to earnings of the year in which stocks were purchased, P/E_t (which assumes perfect forecast of the year's earnings in April), to choose portfolios. Except for the little difference between the third and fourth groups, the portfolios so chosen provided consistently higher holding period returns according to the P/E criterion. Since market participants do not have perfect knowledge of the year's earnings, McEnally used the earnings forecasts of the three most popular advisory services of the P/E ratios. The results were almost the same as those reported for published earnings. Thus the research is consistent with Ball and Brown's study in that financial statement data seems to reflect the economic situation of corporations but either are completely discounted by the market before they are published or do not predict the economic future. In either case, the data are not useful to investors at the time the *SEC* requires disclosure.

B. *Returns to Sophisticated Users of Financial Data*

It may be that the studies cited above approach a complicated problem in too simple a way. It often is claimed that the detailed reports required by the *SEC* are more useful to trained analysts than to the ordinary stockholder. The analyst then passes on his information to his clients or, in any event, trades on the information, thereby bringing its effects to the market. No doubt information about firms does get to the market. But does it get there by means of the financial reports required by the *SEC*?

One way to answer these questions is to examine whether well-trained analysts out-perform the market. I am aware of two studies that test directly the ability of security analysts to use published financial data. John G. Cragg and Burton Malkiel recorded predictions made by five

[10] The geometric mean return for the lowest P/E_{t-1} portfolio is 1.20 compared to the average of all portfolios of 1.15. No significance tests were made. However, McEnally reports that "Simple correlation coefficients between P_t/E_{t-1} and the next year's holding period return for these five holding periods . . . cast doubt on the validity of any relationship" (p. 30).

investment firms of the earnings of 185 corporations whose stock is widely held. In particular, they compared the earnings' growth rate forecast by the analysts with the actual growth rates. They report that "... the remarkable conclusion of the present study is that the careful estimates of security analysts participating in our survey, the bases of which are not limited to public information, perform little better than the past growth rates" [the naive predictor, that the future will be like the past] (p. 83). The second study, by Lyn Pankoff and Robert Virgil, was a controlled, laboratory study. They allowed security analysts to "buy" financial statements of companies in whose stock they can invest. The data are actual data and the stock prices are those that actually prevailed for the stocks whose identity they disguised. While their study is not yet complete, Pankoff and Virgil found that analysts who use financial data (or any other data) do not do as well as they could have had they followed a "naive" buy and hold strategy.

Several indirect tests of the ability of trained analysts to use financial data publicly can be derived from studies of the performance of mutual funds and research departments of brokerage houses. F. E. Brown and Douglas Vickers, William Sharpe, and Michael Jensen studied the performance of mutual funds compared to that of random selections of securities with similar risk characteristics. They used different techniques and all came to the same conclusion. Mutual funds do not earn for their investors a higher rate of return than would have been earned had the investors held a similarly diversified market portfolio, gross of research costs. Nor is the record of the research departments of brokerage houses any better. A study by R. E. Diefenback of the market performance of stocks whose purchase or sale was recommended by twenty-four institutional research services found that their recommendations, if followed, would have yielded returns equivalent to those earned by investments in the Standard and Poor's (S and P) 425 Index. Thus, even mutual fund and brokerage house research department analysts do not benefit from detailed analysis of *SEC* reports, among other data.

Even though the evidence reviewed does indicate that the financial reports required by the *SEC*, when made available, have almost no information content, this does not prove that the required disclosure is not valuable to investors. One might argue that the statements provide a confirmation of data previously released. Because investors know that a corporation's sales, operating expenses, extraordinary gains and losses, assets and liabilities will be reported, they may have some assurance that the preliminary reports, press releases, etc., are not prevarications. Thus when the financial statements are made public the data they contain are fully anticipated. But had it not been for the *SEC*'s disclosure requirements, such a state of affairs might not exist. It is to this consideration that I turn next.

V. An Empirical Analysis of the Effect of the '34 Act on *NYSE* Securities

When the Securities Exchange Act was enacted in June 1934, the United States was in the midst of the Great Depression. Hence, it is difficult to separate the effect of the legislation on the stock market from other economic events, and the effect of the disclosure requirements of the Act from its other provisions. In addition, it is necessary to determine when the legislation affected the stock market, since it might have been anticipated such that, when passed, its impact already had been discounted.

Fortunately, the data and the particular legislative history of the '34 Act allow an

unusual opportunity to test the effect of legislation. Hearings on the '34 Act did not begin until February 1934, nor was there much belief before this date by most observers that such legislation would be enacted (see Sobel). Prior to this time, it was not considered part of the President's legislative "package." Nevertheless, the bill was signed by President Roosevelt in June 1934 and took effect that year, although full compliance did not occur within the year. Therefore, I have considered the period of adjustment to include February 1934 through June 1935. Thus there is a relatively short and distinct period over which the effect of the legislation may be measured.

The effect of the disclosure provisions of the Act may be tested by examining its differential effect on the securities of corporations that were and were not affected by the legislation. At the time of the passage of the Securities Exchange Act of 1934, about 70 percent of stock exchange transactions were made on the *NYSE*, 13 percent on the American (Curb) Exchange, 1.6 percent on the Chicago (Midwest) Stock Exchange, and the balance on nineteen other regional exchanges. The three principal exchanges, at least (the others didn't reply to my inquiries), had similar rules that required listed companies to send certified income statements and balance sheets to stockholders in advance of the annual meeting. The principal reporting requirement imposed by the '34 Act, in addition to the filing of detailed forms, was the required disclosure of sales. Of the 508 corporations whose stock was traded on the *NYSE* in 1934, 193 (38 percent) did not disclose their sales.[11] Since sales are considered very important information by analysts, and the study reported above (Benston (1968)) found sales the only relatively important accounting number, these corporations are considered as those most likely to be affected by the disclosure requirements of the '34 Act.[12]

Thus two samples of *NYSE* corporations can be distinguished: the 314 (62 percent) "disclosure" corporations and 193 (38 percent) "nondisclosure" corporations (with respect to sales). These data allow a fairly comprehensive test of the law, since the '34 Act applied (until 1965) only to corporations whose stock is traded on registered exchanges, and most of these corporations were listed on the *NYSE* in 1934. If disclosure of the sales data required by the '34 Act were meaningful to investors, these effects should be observed in the market returns of the securities affected in the period after the law was effective. As was discussed above in greater detail, if the data disclosed are information, investors would alter their previous estimates of the relative value and/or riskiness of the firms.

The model given by equation (2) above can be used to measure these effects. The $\hat{u}_{jt}$ measure the returns on a security after the effect of changes in the return for the market as a whole is accounted for. A change in the $\hat{u}_{jt}$ represents a revaluation of the present value of future returns from a security. As such, it provides a valuable measure of the information content of mandated disclosure.

The $\hat{\beta}_j$ provide a measure of the relative systematic portfolio risk of a security, its covariance with the market relative to the market's variance.[13] As such the $\hat{\beta}_j$ might be affected if the data disclosed in the financial statements provide information about the risk class of a company and the

[11] Fifty-four percent did not report "cost of goods sold," almost all of whom also did not report sales. (See Table I in Benston (1969b).)

[12] The analysis also was carried out for the corporations that did and did not disclose cost of goods sold, with similar results to those reported.

[13] See Fama, pp. 401–04, for a more complete exposition of this concept.

relationship of its economic value to changes in the economy (as reflected by the market).

The variance of returns from a stock may be stated as:[14]

$$(3) \qquad \sigma^2(r_j) = \beta_j^2\sigma^2(r_M) + \sigma^2(u_j)$$

where the variables are as defined for equation (2) above. The variance of returns on a security, j, can be reduced by reducing the variance of the market, $\sigma^2(r_M)$, the sensitivity of the security to the market, β_j, and/or the residual variance, $\sigma^2(u_j)$. Fisher and Lorie's study shows that the dispersion of the market index is higher for the immediate post-'34 Act period than it was in the years before 1932. It is not until 1945 that it is of a lower magnitude than the late 1920's. A reduction in the β_j, though, would reduce the variance in stock's returns. While a reduction in the variance of the residuals of a stock, $\sigma^2(u_j)$, also reduces the variance, it should be noted that investors can reduce or eliminate this "risk" by holding a portfolio of securities. Let

$$r_p = \sum_{j=1}^{n} \frac{1}{n} r_j$$

be the returns on a portfolio of securities with an equal dollar investment in each security. The variance of the returns on such a portfolio is given by

$$(4) \qquad \sigma^2(r_p) = \left(\sum_{j=1}^{n} \frac{1}{n}\beta_j\right)^2 \sigma^2(r_M) + \sum_{j=1}^{n} \frac{1}{n^2}\sigma^2(u_j) = \bar{\beta}^2\sigma^2(r_M) + \frac{1}{n}\overline{\sigma^2(u)}$$

assuming that $\mathrm{cov}(u_i, u_j)=0$ for $i \neq j$. As n, the number of stocks in the portfolio increases, the residual variance of the portfolio goes to zero (as long as the $\sigma^2(u_j)$ are bounded from above). Nevertheless, the legislators and many economists believe it desirable to reduce the riskiness of individual stocks, since "small investors" cannot purchase enough different securities for efficient diversification.[15]

Thus, three statistics derived from equation (2) are of interest: $\hat{u}_{jt}$; $\hat{\beta}_j$; and $\sigma^2(\hat{u}_{jt})$. Equation (2) was computed for each *NYSE* traded security on monthly data for the period January 1926 through January 1934 (pre-*SEC* period) and July 1935 through November 1941 inclusive (post-*SEC* period). January 1926 was taken as the initial month because this is the first month for which data are available on the Center for Research on Security Prices (*CRSP*) tape of *NYSE* monthly security prices.[16] November 1941 was taken as the last month because it allowed for the longest post-*SEC* period that did not include the next major dislocation—World War II. There are 98 months in the pre-*SEC* period and 90 in the post-*SEC* period. To allow for a sufficient number of observations for the regressions, securities that were traded for less than 10 months in each time period were excluded.[17] This left 466 companies, of whom 290 are disclosure corporations (62 percent) and 176 nondisclosure corporations (38 percent).[18]

Five hypotheses about the effectiveness of required disclosure can be tested with

[14] I am indebted to Michael Jensen for clarifying my thinking on this issue and for the equations presented.

[15] While small investors can purchase mutual funds to obtain diversification, this alternative may not be considered sufficient or even desirable by legislators.

[16] The Center is at the University of Chicago. The tape presently is administered by the Standard Statistics Corporation.

[17] A study of the disclosure practices of corporations who left the *NYSE* before February 1934, and hence were excluded from the analysis, is presented below.

[18] Since the percentages of disclosure and nondisclosure corporations are the same, this procedure does not appear to have introduced any obvious bias.

observations of the $\hat{\beta}_j$ and $\sigma^2(\hat{u}_{jt})$ for the periods before and after disclosure and of the $\hat{u}_{jt}$ for the adjustment period following immediately after required disclosure.

1) Managers avoided disclosure to hide their poor performance:

If the managers did not publish financial data because they wanted to hide their poor performance from their stockholders, the $\hat{u}_{jt}$ of the nondisclosure companies compared to the disclosure companies would be negative since investors would revalue downward the returns to the securities. Similarly, such managers might not disclose financial data to mislead investors about the relative riskiness of the firm, in the hope that stockholders might be willing to hold stock having a lower rate of return if they believed the firm to be less risky than it really was. In this event, disclosure would result in higher $\hat{\beta}_j$ if the market (portfolio) risk of the nondisclosure firms was underestimated and greater $s^2(\hat{u}_{jt})$ if the individual, diversifiable security risk was underestimated and not diversified away.

2) Managers did not disclose because they did not realize the value of the information to investors:

If required disclosure provided investors with valuable information, the $\hat{u}_{jt}$ would change, and the $\hat{\beta}_j$ and $\sigma^2(\hat{u}_{jt})$ might change. However, the direction of the changes cannot be predicted.

3) Required disclosure imposes a cost on corporations without compensating benefits to stockholders:

If managers were disclosing adequately before the legislation was passed (the marginal cost of disclosure equaled its marginal benefits), investors might view disclosure as a net cost imposed on the firm. (Included in this cost is the value of the information to competitors.) In this event, the $\hat{u}_{jt}$ would decrease (as in hypothesis 1)), but there is no reason to believe that the $\hat{\beta}_j$ and $\sigma^2(\hat{u}_{jt})$ would be affected.

4) Required disclosure results in benefits to the market as a whole because investors would prefer stocks on registered exchanges to alternative investments, such as over-the-counter stocks or real estate. However, some costs are imposed on those firms that would not otherwise have disclosed:

If this hypothesis holds, the $\hat{\beta}_j$ and $\sigma^2(\hat{u}_{jt})$ should not be affected, the $\hat{u}_{jt}$ of the firms that were required to disclose might decrease and the $\hat{u}_{jt}$ of firms that were not so affected might increase, as investment in equities traded on the stock exchanges became relatively more attractive to investors. Thus the difference in the $\hat{u}_{jt}$ between disclosure and nondisclosure firms should be positive.

5) Required disclosure did not impose sufficient costs or benefits to be measured:

Should the "null" hypothesis obtain, there would be little change in the $\hat{u}_{jt}$, $\hat{\beta}_{jt}$, or $\sigma^2(\hat{u}_{jt})$. Of course there could be costs to firms not traded at the time (such as the cost of newly registering with the *SEC*) or costs or benefits too small to be measured by the model.

Tests on the $\hat{\beta}_j$ are discussed first because, if the $\hat{\beta}_j$ are stable between periods, the $\hat{u}_{jt}$ can be computed more efficiently by using data from the entire data set, excluding the adjustment periods.

Table 2 gives the mean differences of the $\hat{\beta}_j$ computed for the pre- and post-*SEC* periods, in both algebraic and absolute terms.[19] Distributions of the $\hat{\beta}_j$ show them

[19] The means (and standard errors of the means) of the $\hat{\beta}_j$ in the pre-*SEC* period is .9968 (.0202) for the disclosure corporations and .9854 (.0312) for the nondisclosure group. In the post-*SEC* period, these statistics

to be approximately normally distributed, so the standard errors of the means given in Table 2 can serve as valid summary statistics.[20] Since the stability of the $\hat{\beta}_j$ are dependent, in part, on the number of observations used to compute them, the average number and standard deviations of the underlying observations also are given in Table 2. These data do not indicate any bias between the groups from this source. The algebraic differences between periods for both groups are positive, but hardly greater than zero and almost of the same magnitude and dispersion for the disclosure and nondisclosure groups. The absolute differences are presented because several of the hypotheses do not specify a change in the $\hat{\beta}_j$ of any particular sign, and the algebraic means could mask significant changes. Of course, the absolute means are larger than their algebraic counterparts. Most important, the absolute mean change of the disclosure group is somewhat greater than that of the nondisclosure group, which indicates a smaller change in perceived portfolio riskiness for those corporations affected by the '34 Act. An additional test of this conclusion was made by correlating the $\hat{\beta}_j$ from the pre-*SEC* period with the $\hat{\beta}_j$ from the post-*SEC* period for each group. The correlation coefficient of the pre- and post-*SEC* $\hat{\beta}_j$ (reported in Table 2) is higher for the nondisclosure group, which again indicates a lesser change in the $\hat{\beta}_j$ between periods.

In summary, the tests indicate that the disclosure requirements of the '34 Act had a somewhat lesser effect on the securities of corporations that did not previously disclose sales as compared with those that did. This finding is inconsistent with hypothesis 1) and casts doubt on hypothesis 2).

The stability of the $\hat{\beta}_j$ between periods (or, at the least, a small differential change between the groups) provides support for estimating the $\hat{u}_{jt}$ with the following procedure.[21] Equation (2) was computed for each security that was traded on the *NYSE* for at least five months prior to February 1934 and five months after July 1935, allowing at least ten observations. A total of 487 securities were included under this rule, of which 306 (63 percent) had disclosed their sales and 180 (37 percent) had not prior to the '34 Act. While the

TABLE 2—ESTIMATE OF PORTFOLIO RISK $\hat{\beta}_j$: DIFFERENCES AND CORRELATIONS

POST-*SEC* PERIOD LESS PRE-*SEC* PERIOD[a]

	Disclosure Corporations (290 observations)	Nondisclosure Corporations (170 observations)
Algebraic Differences		
mean	.0320	.0264
standard error of mean	.0218	.0227
Absolute Differences		
mean	.2722	.2133
standard error of mean	.0149	.0159
Correlation of Pre- and Post-Periods	.5725	.7326
Data Underlying β_j Statistics		
average number of observations (months)		
Pre-*SEC* Period	79.0	80.6
Post-*SEC* Period	75.3	76.9
standard deviation of observations (months)		
Pre-*SEC* Period	20.9	22.3
Post-*SEC* Period	10.1	6.9

[a] Pre-*SEC* period: Jan. 1926 through Feb. 1934; Post-*SEC* period: July 1935 through Nov. 1941.

are 1.0115 (.0248) for the disclosure corporations and 1.0006 (.0328) for the nondisclosure companies. The $\hat{\beta}_j$ are lower in both periods for the disclosure group, but insignificantly so.

[20] The standard errors of the means (*SEM*) are relatively small, in part because the sample size (n) is large (SEM = standard derivation/$\sqrt{n-1}$).

[21] The correlation coefficients between the $\hat{\beta}_j$ in the pre- and post-*SEC* periods reported above may not be sufficiently high for some readers to accept this procedure. The alternative procedure would have been to use fewer observations to calculate the statistics, which, I believe, introduces more severe possible biases. In any event, the $\hat{u}_{jt}$ were calculated with $\hat{\beta}_j$'s from the pre-SEC period for the adjustment period and the results were little different from those reported below.

TABLE 3—AVERAGE MONTHLY RESIDUALS IN SUBPERIODS[a]

	Algebraic Means ÷ .001			Absolute Means ÷ .001		
Subperiod	Disclosure	Nondisclosure	Disclosure-Nondisclosure	Disclosure	Nondisclosure	Disclosure-Nondisclosure
2/26 thru 1/28	.074 (.742)	.206 (.986)	−.132 (1.429)	6.653 (.919)	6.459 (.919)	.194 (.774)
2/28 thru 1/30	.020 (.924)	.075 (1.086)	−.056 (1.634)	8.028 (1.604)	7.647 (1.460)	.381 (.748)
2/30 thru 1/32	.026 (1.181)	−.002 (1.525)	−.024 (1.742)	10.475 (2.214)	9.795 (1.913)	.680 (1.007)
2/32 thru 1/34	−.349 (1.638)	−.325 (1.892)	−.024 (2.328)	14.452 (3.707)	12.884 (2.875)	1.568 (1.375)
2/34 thru 6/35 (adjustment period)	.125 (.865)	.728 (.968)	−.603 (1.124)	8.435 (1.115)	7.802 (1.250)	.633 (.812)
7/35 thru 6/37	.227 (.941)	−.263 (1.624)	.491 (1.676)	8.221 (1.813)	7.332 (1.592)	.888 (.728)
7/37 thru 6/39	.017 (.773)	.203 (1.226)	−.186 (1.314)	6.972 (1.185)	6.225 (1.299)	.747 (.609)
7/39 thru 6/41	.077 (.910)	.085 (1.093)	−.008 (1.175)	7.151 (2.961)	6.196 (2.378)	.955 (.743)

[a] Standard deviations shown in parentheses.

number of observations is slightly higher than those used for the $\hat{\beta}_j$ tests, the percentages in the disclosure and nondisclosure groups are almost identical. The $\hat{u}_{jt}$ were computed by running the following "market model" regression for each security

$$r_{jt} = \alpha_j + \beta_j r_{Mt} \tag{5}$$

where t excludes February 1934 through June 1935. The computed $\hat{\alpha}_j$ and $\hat{\beta}_j$ then are used to compute the $\hat{u}_{jt}$ for each month during the "adjustment period," February 1934 through June 1935:

$$\hat{u}_{jt} = \hat{\alpha}_j + \hat{\beta}_j r_{mt} - r_{jt} \tag{5a}$$

where t runs from February 1934 through June 1935. The expected value of the $\hat{u}_{jt}$ is zero in the absence of an unanticipated change in the economic environment, such as the disclosure of previously undisclosed information.

The average (algebraic mean) residuals for each month of the j securities $([\sum_1^n \hat{u}_{jt}]/n$, n = number of securities in each month) were computed for the disclosure and nondisclosure groups. Since several of the hypotheses do not predict a sign of the $\hat{u}_{jt}$, the average absolute residuals for each month $([\sum_1^n |\hat{u}_{jt}|]/n)$ also were computed for each group. Plots of these data were made and summary statistics are given in Table 3 for several subperiods before and after the adjustment period to contrast the behavior of the residuals during that period.

Two basic conclusions can be derived from these data. First, there appears to be somewhat less variance of the residuals of both groups during the months of the adjustment period of February 1934 through June 1935 compared to the following and preceding months. These data indicate that by the time the initial hear-

ings on the Securities Act of 1934 were begun, the impact of the Great Depression on the revaluation of individual shares was largely spent. Second, there is little difference in the behavior of the average residuals of the nondisclosure compared with the disclosure groups. The plots (not presented) show that their residuals behaved almost the same over time, with differences between the two groups being far overshadowed by differences between time periods.

Table 3 is not quite adequate to test the hypothesis that the '34 Act affected each group differently, since the algebraic or absolute mean monthly values and standard deviations over months may not describe the underlying distribution of the individual $\hat{u}_{jt}$. Therefore, Figures 1 and 2 were prepared in which the fractiles (.05 and .95, .10 and .90, .75 and .25, and the median) are plotted for each group over time. If the plot of residuals for one group could have been printed as a transparency and placed over the plot of the other, the reader could see that there is almost no difference in the distribution of residuals between the groups.

The data presented thus far are based on monthly averages of the $\hat{u}_{jt}$. However, it also is necessary to consider the effect of the disclosure statute on individual securities. The possible revaluation of the present value of returns on individual securities need not have taken place entirely in any given month. In addition, it is not clear exactly when corporations made their financial data available as required by the '34 Act. Therefore, the $\hat{u}_{jt}$ were cumulated algebraically for the adjustment period ($\sum_{t=1}^{17} \hat{u}_{jt}$ where t=2/34 through 6/35). As a further check, the $\hat{u}_{jt}$ were cumulated for an additional year ($\sum_{t=1}^{29} \hat{u}_{jt}$ where t=2/34 through 6/36).

Figure 3 presents the data in the form of histograms, in which the distribution of the cumulative residuals for the corporations in the disclosure and nondisclosure group are plotted together. It is clear from these histograms and from the summary statistics (not presented)[22] that there is little difference between the two groups

[22] The algebraic means (and standard deviations) of the cumulative residuals in the February 1934 through June 1935 adjustment period are .0010 (.0307) for the disclosure corporation and .0072 (.0290) for the nondisclosure corporations. In the February 1934 through June 1936 adjustment period, the statistics are −.0004 (.0223) and .0022 (.0206).

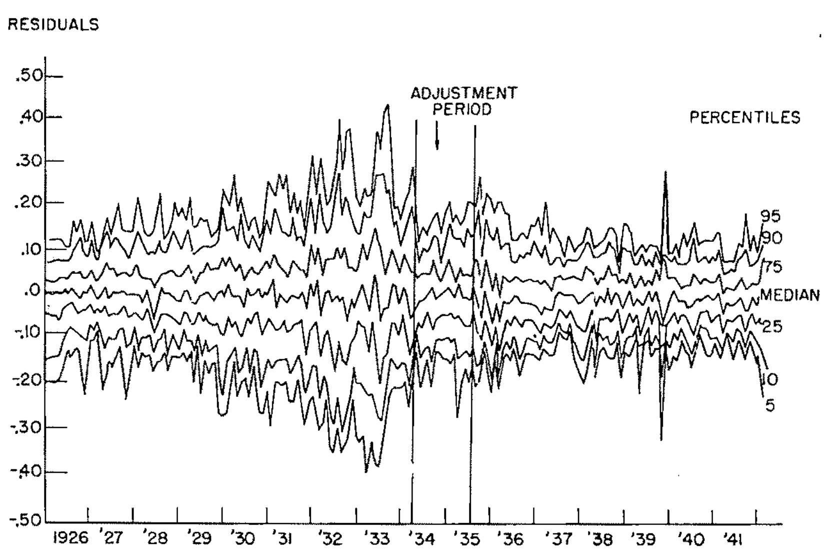


FIGURE 1. NONDISCLOSURE CORPORATIONS DISTRIBUTION OF RESIDUALS ($\hat{u}_{jt}$)

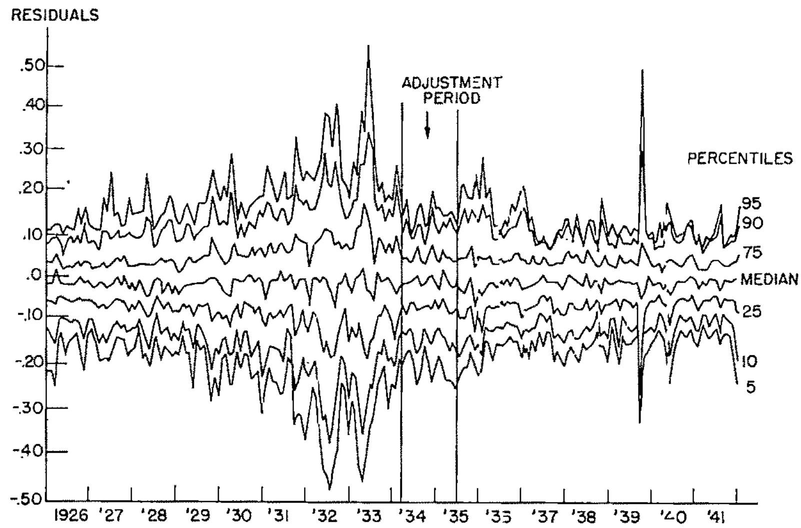


FIGURE 2. DISCLOSURE CORPORATIONS DISTRIBUTION OF RESIDUALS ($\hat{u}_{jt}$)

and that neither group experienced a revaluation of its rates of return that is significantly different from the expected value of zero.

The variance of the residuals of each company, $\sigma^2(\hat{u}_{jt})$ provides an estimate of the volatility of the individual security. A comparison of the mean and distribution of the standard deviations of the residuals of the disclosure and nondisclosure corporations are given in Part A of Table 4. The disclosure corporations show a slightly higher mean standard deviation of residuals in the pre-*SEC* period and, surprisingly, an even greater mean standard deviation in the post-*SEC* period. It is possible, however, that the means of the standard deviations of the $\hat{u}_{jt}$ of each group in the pre- and post-*SEC* periods obscure changes in the variance of individual securities. To test for this possibility, the differences between the standard deviations of the residuals of each company in each period was computed. The mean and distribution of these changes is presented in Part B of Table 4. The mean change is negative for each group, consistent with Part A of the table. Most important, the change in variance of individual stock prices (after accounting for

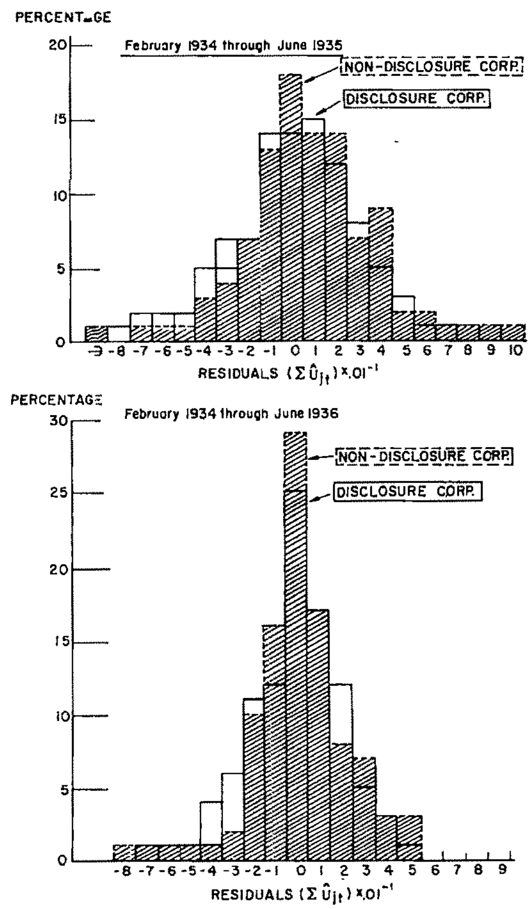


FIGURE 3. HISTOGRAMS OF CUMULATIVE RESIDUALS DISCLOSURE AND NONDISCLOSURE CORPORATIONS

TABLE 4—VARIANCE OF RESIDUALS ($\hat{u}_{jt}$) OF INDIVIDUAL COMPANIES PRE- AND POST-*SEC* PERIODS

	Disclosure	Non-disclosure
A. Means		
Pre-*SEC* Period (1/26–2/34)		
Mean of standard deviations	.1496	.1351
Standard deviation of mean	.0623	.0507
Outlyers		
$>$mean$\pm 1\sigma$	29.1[a]	30.9[a]
$>$mean$\pm 2\sigma$	4.2[a]	3.1[a]
Number of Observations	309	191
Post-*SEC* Period (7/35–11/41)		
Mean of standard deviations	.1033	.0885
Standard Deviation of mean	.0535	.0363
Outlyers		
$>$mean$\pm 1\sigma$	14.9[a]	20.3[a]
$>$mean$\pm 2\sigma$	5.1[a]	2.8[a]
Number of Observations	296	177
B. Change in Variances: Post-*SEC* less Pre-*SEC* Period		
Mean algebraic differences	−.0433	−.0435
Standard deviation of mean	.0534	.0385
Mean absolute differences	−.0540	−.0469
Standard deviation of mean	.0435	.0343
Number of observations	290	176

[a] Shown in percent.

the variance of the stock market as a whole) once the Securities Act of 1934 was effective, is almost the same for those corporations that were and were not affected by the Act.

From these data and the data reported above, I conclude that the '34 Act did not contribute to a reduction in the variance of returns from securities traded on the *NYSE*, as measured by the $\hat{\beta}_j$ (the covariance-variance ratio of security j to the market as a whole) and the standard deviations of the residual returns on securities, $\sigma^2(\hat{u}_{jt})$.[23]

Considering the evidence presented above, I must conclude that the data are consistent only with the hypothesis (5), that the disclosure provisions of the '34 Act were of no apparent value to investors.

VI. Losses by Stockholders

The data used in the tests presented above are of *NYSE* corporations that survived the depression, at least until December 1935 (five months of post-adjustment period data were required for the regressions). A question that should be raised is whether stockholders of corporations that disclosed their sales fared better than those who held stock in nondisclosing corporations. Should this be the case, the findings presented above could be biased because the worst offenders, with respect to disclosure, would have been delisted from the *NYSE*.

To determine whether such was the case, the listing of *NYSE* corporations was traced. The base year chosen was 1929, because this is the last year before the Depression and because I could use a study made by Standard Statistics Company, reported by Lawrence Sloan, pp. 66–74. This study of the income statements of 484 corporations listed on the *NYSE* revealed that 266 (55 percent) reported gross income in 1929 and 218 (45 percent) did not. Table 5 presents the listing history of these securities, showing whether the security was delisted because it went over the counter (*OTC*) or the corporation was merged. Either event is taken to mean a greater than normal loss for stockholders. If this assumption is correct, it is clear that stockholders of corporations that did not disclose gross income in 1929 fared better than those who held stock in the disclosure corporations.

This conclusion is interesting, because it implies that companies that did not disclose their sales were, in fact, better investments than those that did. Such investigations as I could make did not reveal the reasons that some companies disclosed their sales and others did not. A reason consistent with the data would be that those who disclosed had a greater real need

[23] The smaller mean values of the standard deviations of the residuals of securities in both groups in the post-*SEC* compared to the pre-*SEC* period is due primarily to the period of May 1932 through June 1933.

TABLE 5—*NYSE* DELISTING OF CORPORATIONS THAT DID AND DID NOT DISCLOSE GROSS INCOME IN 1929

	Disclosure Corp.			Nondisclosure Corp.		
Number	OTC	Merged	Total	OTC	Merged	Total
Delisted in:						
1929	1	1	2			
1930	1	4	5	1		1
1931	4	2	6	3		3
1932	6	5	11	6		6
1933	5		5	5		5
1934	3		3	1		1
Subtotal: Pre-*SEC*	20	12	32	16		16
1935	4		4	3	1	4
1936	2	1	3			
1937	2	1	3			
Subtotal: 1935–1937	8	2	10	3	1	4
1938	5	1	6	2		2
1939	4	1	5			
1940	2	1	3	1		1
Subtotal: 1938–1940	11	3	14	3		3
Total 1929–1940	39	17	56	22	1	23
On *NYSE* in 1941			210			195
Total			266			218
Percentages						
Delisted 1929–1934	7.5	4.5	12.0	7.3		7.3
Delisted 1934–1937	3.0	.8	3.8	1.4	.4	1.8
Delisted 1938–1940	4.1	1.2	5.3	1.4		1.4
Subtotal	14.6	6.5	21.1	10.1	.4	10.5
Not Delisted as of 1941			78.9			89.5
Total			100.0			100.0

to assure their stockholders of their worth than those who did not.

VII. Investor's Confidence in the Market—Risk and Fairness

A major reason for enactment (and continuance) of the disclosure provisions of the '34 Act was the belief that disclosure was necessary to restore the confidence of investors in the stock market. It is obvious that depressions and the reduction of stock values (either with a "crash" or gradually over a relatively short period) reduce investor confidence and that such events have occurred before and since the passage of the '34 Act and were little affected by it. Therefore, a more meaningful (and charitable) definition of "investor confidence" might be related to the riskiness of returns from securities and to the concept of fairness, that all investors should have equal access to financial information about a company whose shares they own or contemplate buying. Each of these somewhat related concepts is considered in turn.

A reduction of the riskiness of returns, *ceteris paribus*, is considered a benefit because this would reduce the cost of capital to firms and increase investors' confidence in the market. Both assertions are based on the belief that investors are risk averse. (Speculators may prefer risk, but they are

considered by legislators to be a nonpreferred group.) Of course, disclosure, as such, cannot reduce the inherent riskiness of corporations except where disclosure reduces or prevents the risk of fraud. However, disclosure might reduce the risk to the investor of not knowing about significant events (such as a large loss, lawsuit, discovery, etc.) and/or diversifiable risk by reducing the residual variance of security returns (as is discussed above in Section V).

The evidence presented above indicates that disclosure as required by the '34 Act did not reduce fraud, nor did corporations who disclosed their sales fare better in the depression than those who didn't. The data on the variance of the securities of disclosure firms compared to that of nondisclosure firms (Tables 2 and 4), discussed above, indicates that the '34 Act did not have the desired effect on risk (as measured herein). The percentages of large residuals (outlyers) reported in Table 4 also provide evidence on the effectiveness of the '34 Act in reducing the risk of large, presumably unanticipated stock movements. The percentage of outlyers in the pre-*SEC* period was about the same for the disclosure and the nondisclosure firms. However, in the post-*SEC* period, there were relatively fewer mean $\pm 1\sigma$ outlyers for the disclosure group but more mean $\sigma \pm 2$ outlyers for the nondisclosure group.

Some additional evidence on the effect of the '34 Act on the stock market is provided by a study of corporate total capital formation met by stock market issues and publicly sold debt issues compared to private placements in the pre- and post-*SEC* years (through 1954). As I reported in 1969 (see Benston 1969b), the percentage of new issues (net of redemptions) to expenditures on plant and equipment gross and net of price-level adjusted depreciation was the following:[24]

1900–24	30.9 (gross)	58.3 (net)
1919–29	36.8 (gross)	354.3 (net)
1932–38	— (gross)	— (net)
1938–46	6.5 (gross)	51.2 (net)
1946–49	23.1 (gross)	50.2 (net)
1949–54	20.6 (gross)	68.6 (net)

I also found that a ranking of industries by the percentage of private placements to total debt issues was almost exactly the same as a ranking of the extent of bias of the *SEC*'s conservative accounting rules against a fuller reporting of economic events (such as the extent of mining claims, the value of airline routes, etc.). Thus it appears that, contrary to the expectations of its supporters, the '34 Act may have reduced the value of stock markets to corporations and, therefore, to investors.

The concept of fairness is difficult to define operationally. The belief that all investors should have equal access to financial information about a company whose shares they own or contemplate purchasing is perhaps the most important concept, politically, that supports the federal disclosure requirements. While some writers, such as Henry Manne, have argued that investors and the economy are served better when insiders are allowed to profit from information before it is disclosed, legislators argue that all current and potential owners of a corporation have an equal right to information without regard to cost. But since any information must be available to someone before it is known to all, this is a nonoperational concept in its extreme form. And, insofar as the *SEC*'s disclosure requirements require the publication of useless or untimely data (as seems to be the case), the '34 Act has not served its purpose.

Nevertheless, the stock market could be considered "fair" if the prices of securities at any point in time are unbiased estima-

[24] Stock redemptions exceeded new issues in the 1932–1938 period.

tors of their intrinsic values, at least with respect to the financial data which corporations must disclose under the '34 Act. Then whenever an investor decides to buy or sell or hold a security, he can be assured that the market price has discounted completely the financial information. The average investor need not worry about discovering some important financial information about which he is unaware. He will just as often find himself buying or selling a security that is "overvalued" as "undervalued." In this event, the market would be "efficient" in what Fama calls the semi-strong form of the martingale hypothesis.

Fama reviews the theory and evidence on efficient capital markets and concludes that ". . . for the purposes of most investors the efficient markets model seems a good first (and second) approximation to reality. In short, the evidence in support of the efficient markets model is extensive, and (somewhat uniquely in economics) contradictory evidence is sparse" (p. 416). But this evidence is based on data from years after enactment of the '34 Act. The Act may have altered the way in which information gets to the market and the speed with which it is dispersed such that a previously inefficient market became efficient.

The data presented above, that the disclosure required by the '34 Act had no measurable effect on the residual market prices of companies that did and did not disclose their sales, are consistent with the hypothesis that the market was efficient before the legislation was enacted, at least with respect to the financial data.[25] In addition, runs tests on the signs of price changes before 1934 of the securities of the disclosure and nondisclosure corporations revealed that the price changes of both groups confirmed to a random walk.[26] The results of this weaker test of the efficient markets hypothesis is consistent with the belief that the '34 Act did not make the stock market a "fairer game" for investors.

VIII. Summary and Conclusions

The Securities Exchange Act of 1934 is called a "disclosure statute" because its purpose is to force corporations to provide the public with financial data as prescribed by the *SEC*. The Act was passed and extended because of the belief that such disclosure was required to correct the abuses of the pre-Great Depression period, provide information necessary for investors to allocate their resources wisely and efficiently, and make the stock market a "fair game" for the average investor.

A review of the evidence on fraudulently or misleadingly prepared financial statements prior to enactment of the '34 Act revealed very little evidence of abuses in reporting. The assertion by the *SEC* that "pools" and other presumably manipulative devices were made possible by inadequately disclosed financial information is not supported by the data. Proportionately, there was little difference between the reporting practices of corporations whose securities were subject to pools as those who were not.

The value of reported financial data for investors' decisions is based on the as-

[25] These data also are consistent with the hypothesis that published financial data have no information content.

[26] The mean percentage deviations between actual and expected (assuming a random distribution) total runs in the pre-*SEC* and post-*SEC* periods are −1.2 and 4.9 percent for the disclosure corporations and −1.5 and 7.5 percent for the nondisclosure corporations. On a probability basis, the percentage of individual stocks whose runs were significantly different at the 5 percent level from expected are 4.3 and 7.8 for the pre- and post-*SEC* disclosure corporations and 2.3 and 12.2 for the pre- and post-*SEC* nondisclosure corporations. It appears that the security prices of both groups more nearly conformed to a random walk in the pre-*SEC* period than in the post-*SEC* period. Most interesting for present purposes, prices of the securities of corporations that did disclose appear to reflect a less efficient market after the '34 Act became effective than did those of corporations that did not disclose their sales.

sumption that the data the *SEC* requires be made public is useful and timely. The *SEC*'s adherence to historically based, "conservative" accounting procedures reduces the value of the numbers. Nevertheless, the question is an empirical one for which an empirical answer is sought. The extant statistical studies that relate published accounting statement data with stock prices all lead to the conclusion that the data either are not useful or have been fully impounded into stock prices before they are published.[27] Since these studies use relatively simple decision models, evidence on the ability of professional analysts to use financial data for stock choices was reviewed. This evidence supports the conclusion that the accounting statements either are not useful, timely, or both.

However, these findings are based on data from the post-*SEC* period. It may be that the *SEC* has created a climate of confidence in financial data such that the public can accept information as it becomes available during the year. In this event, the annual financial statements, when published, simply confirm that which was previously released, and therefore do not affect stock prices. Investors could believe the information because they know that a reputable government agency, the *SEC*, is concerned about the veracity of financial statements.

The hypothesis was tested by examining statistically the change in riskiness and returns in the stock prices on the *NYSE* before, after, and as a result of the Securities Exchange Act of 1934. Financial statements were available for almost all *NYSE* corporations, but only 62 percent reported their sales or other similarly important information. These corporations are a control sample against which the effects of the '34 Act can be measured. The Sharpe-Markowitz "market model" was used to account for general stock market changes and to provide estimates of the market risk of stocks for each of the samples (disclosure and nondisclosure corporations) in the periods before and after the regulations became effective. Revaluations of securities as a result of the required disclosure of information also were measured with the model. All of the many measurements and analyses show that the '34 Act's financial disclosure requirements had no measurable effect on the securities of the corporation presumedly affected.

The effect of the '34 Act on investors' confidence and on the fairness of the market also was analyzed. Riskiness of securities, as measured by the variance of market prices net of covariance with the market, does not seem to have been reduced by the Act. Nor were the relative percentage of large price movements reduced. Also, the effect of the '34 Act on the capital market may have been perverse, since the percentage of corporate expenditures on plant and equipment financed with new public stock and debt issues was lower in the post- than in the pre-*SEC* decades. With respect to fairness, the evidence that the disclosure requirements did not result in a revaluation of the securities of the affected firms and the conformity of price changes before 1934 to a random walk indicates that the pre-*SEC* stock market was a fair game for investors.

The conclusion of this study, then, must be that the disclosure requirements of the Securities Exchange Act of 1934 had no measurable positive effect on the securities traded on the *NYSE*. There appears to have been little basis for the legislation and no evidence that it was needed or desirable. Certainly there is doubt that more required disclosure is warranted.

[27] An exception may be Beaver's study of preliminary reports.

REFERENCES

R. Ball and P. Brown, "An Empirical Evaluation of Accounting Income Numbers," *J. Accounting Res.*, autumn 1968, *6*, 159–78.

W. H. Beaver, "The Information Content of Annual Earnings Announcements," *J. Accounting Res., Empirical Research in Accounting: Selected Studies 1968,* 67–92.

G. J. Benston, "Published Corporate Accounting Data and Stock Prices," *J. Accounting Res., Empirical Research in Accounting: Selected Studies 1967,* 1–54.

———, (1969a) "The Effectiveness and Effects of the SEC's Accounting Disclosure Requirements," in H. G. Manne, ed., *Economic Policy and the Regulation of Corporate Securities,* Washington 1969.

———, (1969b) "The Value of the SEC's Accounting Disclosure Requirements," *Accounting Rev.,* July 1969, *44,* 515–32.

M. Blume, "The Assessment of Portfolio Performance," unpublished doctoral dissertation, Univ. Chicago 1968.

———, "Portfolio Theory: A Step Towards its Practical Application," *J. Bus. Univ. Chicago,* Apr. 1970, *43,* 152–73.

F. E. Brown and D. Vickers, in I. Friend et al., *A Study of Mutual Funds,* Washington 1962, ch. 5, 289–358.

J. G. Cragg and B. G. Malkiel, "The Consensus and Accuracy of Some Predictions of the Growth of Corporate Earnings," *J. Finance,* Mar. 1968, *22,* 67–84.

R. F. DeBedts, *The New Deal's SEC,* New York 1964.

R. E. Diefenbach, "How Good is Institutional Research?" *Financial Analysts J.,* Jan./Feb. 1972, *28,* 54–60.

E. Fama, "Efficient Capital Markets; A Review of Theory and Empirical Work," *J. Finance,* May 1970, *25,* 383–417.

———, L. Fisher, M. Jensen, and R. Roll, "The Adjustment of Stock Prices to New Information," *Int. Econ. Rev.,* Feb. 1969, *10,* 1–21.

L. Fisher and J. H. Dorie, "Some Studies of Variability of Returns on Investments in Common Stocks," *J. Bus. Univ. Chicago,* Apr. 1970, *43,* 99–134.

I. Friend and E. Herman, "The SEC Through a Glass Darkly," *J. Bus. Univ. Chicago,* Oct. 1964, *37,* 382–405.

R. Hagerman, "An Investigation of Effects of Disclosure and Audits on the Efficiency of the Market for Bank Stock," unpublished doctoral dissertation, Univ. Rochester 1972.

M. C. Jensen, "The Performance of Mutual Funds in the Period 1954–64," *J. Finance,* May 1968, *23,* 389–416.

H. G. Manne, *Insider Trading and the Stock Market,* New York 1966.

H. M. Markowitz, *Portfolio Selection: Efficient Diversification of Investments,* New York 1959.

R. McEnally. "Information Effect and P/E Ratios," *Miss. Vall. J. Bus. Econ.,* spring 1970, *5,* 24–38.

L. D. Pankoff and R. Virgil, "Some Preliminary Findings from a Laboratory Experiment on the Usefulness of Financial Accounting Information to Security Analysts," *J. Accounting Res., Empirical Research in Accounting: Selected Studies 1970.*

W. D. Rich, *Legal Responsibility and Rights of Public Accountants,* New York 1935.

W. Z. Ripley, *Trusts, Pools and Corporations,* rev. ed., New York 1916.

———, *Main Street and Wall Street,* Boston 1927.

W. F. Sharpe, "Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk," *The Journal of Finance,* Sept. 1964, *19,* 425–42.

L. H. Sloan, *Everyman and His Common Stocks,* New York 1931.

R. Sobel, *The Big Board: A History of the New York Stock Market,* New York 1965.

G. J. Stigler, (1964a) "Public Regulation of the Securities Markets," *J. Bus. Univ. Chicago,* Apr. 1964, *37,* 117–42.

———, (1964b) "Comment," *J. Bus. Univ. Chicago,* Oct. 1964, *37,* 414–22.

F. M. Wheat, "Truth in Securities Three Decades Later," *Howard Law Review,* Winter 1967, *13,* 100–107.

———, Wheat Report, *Disclosure to Investors, A Reappraisal of Administrative Policies Under the '33 and '34 Securities Acts,* Report and Recommendations to the Securities and Exchange Commission from the Disclosure Policy Committee, Mar. 1969.

Landell vs. Lybrand, 254, Pa. 496, 107 Atl. 783, 1919.

Moody's Investors Services, *Manual of In-*

vestments, American and Foreign, Industrial Securities, New York, various years.
Securities and Exchange Commission, *The Work of the Securities and Exchange Commission,* Washington, May 1967.
———, *A 25 Year Summary of the Activities of the Securities and Exchange Commission,* Washington 1959.
Standard and Poors, Industrial 425 Index, New York, various years.
Ultramares Corp. v. Touche, Niven & Co., 229 App. Div. 581, 243 N.Y. Supp. (1st Dept. 1930), rev., 225 N.Y. 170, 174 N.E. 441, 1931.
U.S. Congress, Senate Committee on Banking & Currency, Stock Exchange Practices, *Hearings,* on S. Res. 84 and S. Res. 56, 73 Cong., Part 17, Washington 1934.

Optimal Open Market Strategy: The Use of Information Variables

By John H. Kareken, Thomas Muench, and Neil Wallace*

Undeniably, the Federal Open Market Committee (*FOMC*), which decides and executes open market policy for the Federal Reserve System, is much of the time in the dark about what has been going on in the economy. In saying this, we do not mean to be at all derisive. It is simply that the Committee decides policy each day. And yet, most days it does not know the values for the preceding day of all the various economic variables—or, indeed, even the obviously important ones. Thus, only on the first day of the statement week does the Committee know the value of the money stock for the day before. And on how many days of, for example, the calendar quarter does it know, to a reasonable approximation, the value of *GNP* for the day before? Certainly not many.

We can imagine it not mattering in the slightest that the *FOMC* is, as we have said, much of the time in the dark. The consensus judgment of economists is surely, though, that the past influences the present. And information lags being what they are, the presumption must be that most days there are initial conditions which the Committee does not know. So it would seem not entirely idle to investigate how a central bank, or the responsible decision unit thereof, should conduct open market operations if much of the time it is in the dark or, in other words, if most days there are some unknown initial conditions. That is what we have done, although using some simplifying assumptions. In Sections I and II of this paper we present an open market operations strategy for the central bank which much of the time has to contend with unknown initial conditions. Then, in Section III, we compare this strategy with one which we believe is not unlike that of the *FOMC*.

The strategy of Sections I–II is optimal, but so far as we know only on our simplifying assumptions. There are three assumptions which, because they are so unrealistic, we particularly regret having had to make. The first is that the central bank has but one target variable. The second is that the economic structure, although stochastic, is linear. And the third is that the central bank knows with certainty the constants and coefficients of this structure. These assumptions, taken together, imply that the central bank has to optimize only over the current day, not a succession of days extending far (perhaps indefinitely far) into the future.[1] That is why they are so convenient. Still, we should be happier if we had been able to manage with different assumptions—among them, that the central bank is uncertain about the constants and coefficients of the structure constraining it.[2]

What we think of as our information as-

* Staff members of the department of economics, University of Minnesota, and consultants to the Federal Reserve Bank of Minneapolis. All views expressed herein are the sole responsibility of the authors and should not be interpreted as representing those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

[1] For us, the "day," an arbitrary but brief stretch of time, is the decision period. There is also the "week," made up of several days, and the "quarter," made up of several weeks.

[2] In the concluding section of the paper, we argue that this is what for realism's sake we should have assumed. It might seem ridiculous to assume that the central bank knows the economic structure which determines, say, calendar-day *GNP*. But if sufficient uncertainty is assumed, this is not so.

sumptions seem to us, however, to be quite realistic. In the economic structure of this paper, which holds over days, there are four variables: a nominal rate of interest; two monetary aggregates, the central bank's portfolio of assets and the stock of deposits; and nominal *GNP*, which we take as the central bank's target variable. And, as we assume, both the interest rate and its asset portfolio are observed by the central bank essentially without lapse of time. At the beginning of each day, it finds out what the values of the interest rate and its asset portfolio were for the day before. (It obviously has to know the value of either the interest rate or its portfolio, since for us it must have used one or the other of these variables as its instrument variable.) But neither the deposit stock nor *GNP* is always observed without lapse of time. The central bank gets observations on the daily deposit stock only at the beginning of each week and observations on daily *GNP* only at the beginning of each quarter. Consequently, only on the first day of the quarter can it determine initial conditions exactly. On all the other days, it cannot. On all the days of the quarter except the first, it must therefore guess initial conditions.

What we show below is that it is optimal for the central bank, in guessing initial conditions for the current day, to use all such observations as have become available since it last decided open market policy. It should always use the most recent observation on the interest rate or its portfolio. And on the first day of the week it should also use the newly received observations on the daily money stock. (Of course, on the first day of the quarter it should use the newly received observations on daily *GNP*.) Precisely how it should use these observations—or how optimal open market policy depends on these observations—is what we explain in some detail in Sections I and II.

We refer to all those variables which the optimizing central bank uses in guessing initial conditions—or, alternatively, in deciding open market policy—as information variables. This is the natural phrase, since observations on these variables constitute information about initial conditions. It would be misleading to refer to any one of them as an intermediate target variable, although this is an oft-heard phrase. For, as we show in Section III, unless the economic structure is very special, the optimizing central bank does not use an intermediate target variable. It does not, that is, start out by calculating a target value or sequence of values for the deposit stock or for any other variable, except of course its true or ultimate target variables (for us, nominal *GNP*), and then strive over several days or weeks for equality between target and actual values. It does, in effect, calculate sequences of values—expected values, however, not target values. Indeed, it does this each day, but only so that subsequently it can compare certain of these values with actual values. This it has to do in deciding open market policy, for comparisons of actual and previously expected values are what determine its guesses about initial conditions. But once having observed discrepancies, it has to calculate new sequences of values—expected values, as we think of them. If there are discrepancies, then initial conditions are not what they were expected to be. Expected values are therefore not what they were expected to be.

Admittedly, there is little point in fussing about descriptive phrases. It is perhaps acceptable to go on using the customary one—intermediate target variable. Those who do should, though, keep in mind that the optimizing central bank likely changes its target value or values each day. It has to make a change, at least on days when it observes discrepancies between actual and previously calculated target (for us, ex-

pected) values. Once discrepancies have been observed, the value of, say, the deposit stock consistent with the desired value of GNP, or any other ultimate target variable, is not what it was. It cannot be, for initial conditions are not as expected.

I. The Optimal Open Market Strategy

In this section, we first set out those simplifying assumptions to which we have referred and then show how, based on these assumptions, the central bank should proceed day by day in the conduct of open market operations. As it turns out, what the central bank should do each day depends on certain expectations. It is in Section II, however, that we provide the formulae for these expectations.

We begin by giving some market equilibrium conditions. The first, for the goods market, is

$$(1) \quad Y_t = \alpha_0 + \alpha_1 R_t + a_t(1), \quad \alpha_1 < 0$$

where Y_t is nominal GNP, R_t is the nominal rate of interest, and $a_t(1)$ is a random disturbance, all for day t. The second equilibrium condition, for the money market, is

$$(2) \quad R_t = \beta_0 + \beta_1 Y_t + \beta_2 P_t + \beta_3 [a_t(2) - a_t(3)]$$

where P_t is the central bank's asset portfolio and $a_t(2)$ and $a_t(3)$ are random disturbances, all for day t.

Underlying the second equilibrium condition are two private-sector structural (demand) equations. One is

$$(3) \quad \begin{aligned} M_t^d &= \alpha_2 + \alpha_3 Y_t + \alpha_4 R_t + a_t(2), \\ &\alpha_3 > 0, \quad \alpha_4 < 0 \end{aligned}$$

where M_t^d is the private sector's desired stock of deposits for day t or, since by assumption there is no coin or currency, its desired stock of money. The other, already having been transformed, is

$$(4) \quad M_t = \alpha_5 + \alpha_6 R_t + \alpha_7 B_t^d + a_t(3), \ \alpha_6, \alpha_7 > 0$$

where M_t is the banking sector's day t stock of deposit liabilities and B_t^d is its desired stock of reserves for day t. The second equilibrium condition follows then from $M_t^d = M_t$ and $P_t = B_t^d$.[3] And, as is easily established,

$$\beta_0 = \frac{\alpha_5 - \alpha_5}{\alpha_6 - \alpha_4}$$

$$\beta_1 = \frac{\alpha_3}{\alpha_6 - \alpha_4} > 0$$

$$\beta_2 = \frac{-\alpha_7}{\alpha_6 - \alpha_4} < 0$$

and

$$\beta_3 = \frac{1}{\alpha_6 - \alpha_4} > 0$$

The central bank is, by assumption, subjectively certain about the α_i. It is not, however, about the $a_t(i)$, which as noted above are random variables. All the central bank knows about the $a_t(i)$ are the stochastic processes which generate them. These are

$$(5) \quad a_t(i) = \rho_i a_{t-1}(i) + u_t(i), \quad |\rho_i| < 1$$

where the $u_t(i)$, random variables, are normally distributed, with

$$E_t u_t(i) = 0$$

and

$$E_t[u_t(i) u_{t+h}(j)] = \begin{cases} V_t u_t(i) = V u(i) & \text{for } i = j \text{ and } h = 0 \\ 0 & \text{otherwise} \end{cases}$$

The symbol $E_t(\cdot)$, which appears very often on the pages following, stands for the expected value of $(\cdot)$ as of the beginning of day t, or for the conditional expectation of $(\cdot)$, which depends on all observations through day t−1.

[3] In equating P_t and B_t^d, we ignore a number of so-called determinants of bank reserves (for example, float).

The past does then influence the present, although in a particularly uncomplicated way. But it is not important that the assumed lag structure is simple. If it were not, there would still be an optimal strategy and its derivation would be essentially the same as that which follows. It is only required that the economic structure be linear.[4]

The central bank's loss function is

$$L = \sum_{1}^{Q} (Y_t - \tilde{Y})^2$$

where Q is the number of days in the quarter and $\tilde{Y}$ is the desired or target value of Y_t.[5] It minimizes expected loss, but subject to equations (1), (2), and (5) and of course the information at hand, which as already explained is most of the time incomplete.

To be more precise, what the central bank minimizes at the beginning of day 1 is E_1L. And since, for $j>0$, $E_{i-j}(\cdot) = E_{t-j}[E_t(\cdot)]$,

$$E_1L = E_1[(Y_1 - \tilde{Y})^2 + E_2[(Y_2 - \tilde{Y})^2 + \dots + E_{Q-1}[(Y_{Q-1} - \tilde{Y})^2 + E_Q(Y_Q - \tilde{Y})^2]]\dots]$$

The central bank minimizes E_1L by first minimizing $E_Q(Y_Q - \tilde{Y})^2$. The minimum of $E_Q(Y_Q - \tilde{Y})^2$, denoted by $\tilde{E}_Q(Y_Q - \tilde{Y})^2$, is in general a function of all initial conditions. Then the central bank minimizes

$$E_{Q-1}[(Y_{Q-1} - \tilde{Y})^2 + \tilde{E}_Q(Y_Q - \tilde{Y})^2]$$

And so it proceeds, until in the end it has minimized

$$E_1[(Y_1 - \tilde{Y})^2 + \tilde{E}_2[(Y_2 - \tilde{Y})^2 + \dots + \tilde{E}_Q[(Y_Q - \tilde{Y})^2]]\dots]$$

That the central bank, proceeding in the way described, does indeed minimize E_1L is established by induction.[6] Whatever the outcomes for days 1, 2, . . . , Q−1 or the observations at hand at the beginning of day Q,

$$\tilde{E}_Q(Y_Q - \tilde{Y})^2 \leq \overline{E}_Q(Y_Q - \tilde{Y})^2$$

where $\overline{E}_Q$ is the expected loss associated with any policy whatever for day Q.

We must now establish that

if (a) $\tilde{E}_{t+1}(\sim) \leq \overline{E}_{t+1}(-)$

where

$$\tilde{E}_{t+1}(\sim) = \tilde{E}_{t+1}[(Y_{t+1} - \tilde{Y})^2 + \tilde{E}_{t+2}[(Y_{t+2} - \tilde{Y})^2 + \dots + \dots \tilde{E}_Q(Y_Q - \tilde{Y})^2]]\dots]$$

and

$$\overline{E}_{t+1}(-) = \overline{E}_{t+1}[(Y_{t+1} - \tilde{Y})^2 + \overline{E}_{t+2}[(Y_{t+2} - \tilde{Y})^2 + \dots + \overline{E}_Q(Y_Q - \tilde{Y})^2]]\dots],$$

then (b) $\tilde{E}_t(\sim) \leq \overline{E}_t(-)$

By hypothesis (inequality (a)) and the nature of the expectations operator,

$$\overline{E}_t[\tilde{E}_{t+1}(\sim)] \leq \overline{E}_t[\overline{E}_{t+1}(-)]$$

Adding $\overline{E}_t(Y_t - \tilde{Y})^2$ to both sides,

$$\overline{E}_t[(Y_t - \tilde{Y})^2 + \tilde{E}_{t+1}(\sim)] \leq \overline{E}_t(-)$$

The desired conclusion (inequality (b)) follows, since by definition $\tilde{E}_t(\sim)$ is less

[4] Nor do the stochastic processes have to be first-order Markovian. Incidentally, it is easy enough to do away with the influence of the past. One has only to assume, for all i, that $\rho_i = 0$.

[5] We might have assumed the central bank's loss function to be

$$L = \left(\sum_{1}^{Q} Y_t - \sum_{1}^{Q} \tilde{Y}\right)^2$$

where $\Sigma_1^Q Y_t$ is the familiar quarterly *GNP*. It seemed unreasonable, though, to force the central bank only to "average" the Y_t or, in effect, to make it not care about within-quarter fluctuations of Y_t. The assumption of a constant target value, $\tilde{Y}$, is of no consequence. We could easily have managed with an arbitrary path of target values, $\tilde{Y}_t$.

[6] Those familiar with dynamic programming will recognize the following proof.

than or equal to the left-hand side of this inequality.

The central bank has two possible instrument variables, however, P, its asset portfolio and R, the nominal interest rate. It therefore does not simply minimize expected loss by choice of a value for a given instrument variable. Rather, it minimizes by choice of a value for R and a value for P and then by comparing its two minimum expected losses.

With R as the instrument variable, the reduced-form equation for Y_t is equation (1) or, by equations (5),

$$Y_t = \alpha_0 + \alpha_1 R_t + \rho_1 a_{t-1}(1) + u_t(1) \tag{6}$$

By definition, though, $a_{t-1}(i)$ is the sum of $E_t a_{t-1}(i)$, the expectation of $a_{t-1}(i)$ as of the beginning of day t, and a forecast error $e_t(i)$, the deviation of $a_{t-1}(i)$ from that expectation; that is

$$a_{t-1}(i) = E_t a_{t-1}(i) + e_t(i) \tag{7}$$

Therefore,

$$\begin{aligned} Y_t = \alpha_0 + \alpha_1 R_t + \rho_1 E_t a_{t-1}(1) \\ + \rho_1 e_t(1) + u_t(1) \end{aligned} \tag{8}$$

And so

$$\begin{aligned} &E_Q(Y_Q - \bar{Y})^2 \\ &= [\bar{Y} - \alpha_0 - \alpha_1 R_Q - \rho_1 E_Q a_{Q-1}(1)]^2 \\ &\quad + E_Q[\rho_1 e_Q(1)^2 + u_Q(1)^2] \end{aligned}$$

Since the second of the terms on the right-hand side of the equation does not depend on R_Q, the optimal or minimizing value of R_Q satisfies the equation

$$\bar{Y} = \alpha_0 + \alpha_1 R_Q + \rho_1 E_Q a_{Q-1}(1) \tag{9}$$

Also,

$$\begin{aligned} \tilde{E}_Q(Y_Q - \bar{Y})^2 &= \rho_1^2 E_Q e_Q(1)^2 + E_Q u_Q(1)^2 \\ &= \rho_1^2 \Pi_Q(1, 1) + \Phi(1, 1) \end{aligned}$$

where Φ is the variance-covariance matrix of the $u_t(i)$ and Π_t is the variance-covariance matrix of the $e_t(i)$.[7]

[7] That is, $\Pi_t = E_t(e_t e_t')$ where $e_t' = [e_t(1), e_t(2), e_t(3)]$.

With P as the instrument variable, the reduced-form equation for Y_t is

$$\begin{aligned} Y_t = \delta_0 + \delta_1 P_t + \delta_2 a_t(1) \\ + \delta_3[a_t(2) - a_t(3)] \end{aligned} \tag{10}$$

or, by equations (5),

$$Y_t = \delta_0 + \delta_1 P_t + \Delta_\rho' a_{t-1} + \Delta' u_t \tag{11}$$

where

$$\begin{aligned} \Delta' &= (\delta_2, \delta_3, -\delta_3) \\ \Delta_\rho' &= (\rho_1\delta_2, \rho_2\delta_3, -\rho_3\delta_3) \\ a_t' &= [a_t(1), a_t(2), a_t(3)] \\ u_t' &= [u_t(1), u_t(2), u_t(3)] \\ \delta_0 &= \frac{\alpha_0 + \alpha_1\beta_0}{1 - \alpha_1\beta_1} \\ \delta_1 &= \frac{\alpha_1\beta_2}{1 - \alpha_1\beta_1} > 0 \\ \delta_2 &= \frac{1}{1 - \alpha_1\beta_1} > 0 \end{aligned}$$

and

$$\delta_3 = \frac{\alpha_1\beta_3}{1 - \alpha_1\beta_1} < 0$$

Then, by equations (7),

$$\begin{aligned} E_Q(Y_Q - \bar{Y})^2 &= (\bar{Y} - \delta_0 - \delta_1 P_Q - \Delta_\rho' E_Q a_{Q-1})^2 \\ &\quad + E_Q(\Delta' u_Q u_Q' \Delta + \Delta_\rho' e_Q e_Q' \Delta_\rho) \end{aligned}$$

It follows that the optimal value of P_Q satisfies the equation

$$\bar{Y} = \delta_0 + \delta_1 P_Q + \Delta_\rho' E_Q a_{Q-1} \tag{12}$$

and

$$\tilde{E}_Q(Y_Q - \bar{Y})^2 = \Delta'\Phi\Delta + \Delta_\rho'\Pi_Q\Delta_\rho$$

The central bank therefore uses R or P as its instrument variable for day Q, depending on whether

$$\begin{aligned} &\Phi(1, 1) + \rho_1^2 \Pi_Q(1, 1) \\ &\qquad \lesseqgtr \Delta'\Phi\Delta + \Delta_\rho'\Pi_Q\Delta_\rho \end{aligned} \tag{13}$$

If it uses R, then open market policy is

determined by equation (9). And if it uses P, then open market policy is determined by equation (12). It follows that $E_{Q-1}[(Y_{Q-1}-\bar{Y})^2+\tilde{E}_Q(Y_Q-\bar{Y})^2]$ is either

$$E_{Q-1}[(Y_{Q-1} - \bar{Y})^2 + \Phi(1, 1) + \rho_1^2\Pi_Q(1, 1)]$$

or

$$E_{Q-1}[(Y_{Q-1} - \bar{Y})^2 + \Delta'\Phi\Delta + \Delta_\rho'\Pi_Q\Delta_\rho]$$

But Φ is a constant. And as we show below (Section II), so is Π_t; that is to say, for all t, it is independent of the sequence of instrument variable choices and values up through day t−1. Consequently, the day Q−1 choices of instrument variable and value can affect only $E_{Q-1}(Y_{Q-1}-\bar{Y})^2$ and thus should be made so as to minimize this term. The minimization problem of day Q−1 is therefore exactly the same as that of day Q. The day Q−1 analogue of equation (9) therefore determines the optimal value of R_{Q-1}. The day Q−1 analogue of equation (12) determines the optimal value of P_{Q-1}. And the analogue of inequalities (13) determines whether R or P is the instrument variable for day Q−1.

We might continue through the minimization process, but likely it has already become apparent that, with Π_t independent of past instrument variable choices and values, there is only one term of

$$E_t[(Y_t - \bar{Y})^2 + \tilde{E}_{t+1}[(Y_{t+1} - \bar{Y})^2 + \dots + \tilde{E}_Q(Y_Q - \bar{Y})^2]] \dots],$$

namely $E_t(Y_t-\bar{Y})^2$, that depends on policy for day t. The optimal value of R_t, denoted by $\tilde{R}_t$, is therefore the solution of the day t analogue of equation (9). That is

$$\text{(14)} \quad \tilde{R}_t = [\bar{Y} - \alpha_0 - \rho_1 E_t a_{t-1}(1)]/\alpha_1$$

And the optimal value of P_t, denoted by $\tilde{P}_t$, is the solution of the day t analogue of equation (12),

$$\text{(15)} \quad \tilde{P}_t = (\bar{Y} - \delta_0 - \Delta_\rho' E_t a_{t-1})/\delta_1$$

Finally, the central bank uses R or P as its instrument variable, depending on whether

$$\text{(16)} \quad \Phi(1, 1)+\rho_1^2\Pi_t(1, 1) \lesseqgtr \Delta'\Phi\Delta+\Delta_\rho'\Pi_t\Delta_\rho$$

If the central bank always had complete information, then $\Pi_t=0$ would hold for all t. And it would use R or P as its instrument variable, depending on whether[8]

$$\Phi(1, 1) \lesseqgtr \Delta'\Phi\Delta$$

There is therefore an easy interpretation of the second terms on both the left- and right-hand sides of inequality (16). With R as the instrument variable, there is some additional expected loss, $\rho^2\Pi_t(1, 1)$, attributable to the central bank not having complete information or to the existence of our assumed information lags. And with P as the instrument variable, there is also some additional expected loss, $\Delta_\rho'\Pi_t\Delta_\rho$, attributable to the central bank not having complete information.

II. The Central Bank's Expectations and a Proof of Independence

To decide open market policy for day t, the central bank has to know $E_t a_{t-1}$ and Π_t. In this section, we therefore derive formulae for these expectations and then, making use of the formula for Π_t, give an inductive proof that Π_t is independent of past policy decisions.

As is well known, if Y_{t-1} and $X_{t-1}(1)$, $X_{t-1}(2), \dots, X_{t-1}(N)$ are normal random variables and the $X_{t-1}(i)$ the only variables observed at the beginning of day t are uncorrelated with each other, then the expectation of Y_{t-1} with the $X_{t-1}(i)$ known, or as of the beginning of day t, is

$$\text{(17)} \quad E_t Y_{t-1} = E_{t-1} Y_{t-1} + \sum_i \eta_t(i)[X_{t-1}(i) - E_{t-1}X_{t-1}(i)]$$

[8] If $\rho_i=0$ for all i, then the central bank would effectively always have complete information, even with information lags, and it would use this inequality in making its instrument variable choice. This inequality is essentially like those given by William Poole, p. 206, and Kareken, p. 389.

$E_{t-1}Y_{t-1}$ is the expectation of Y_{t-1} with the $X_{t-1}(i)$ unknown, or as of the beginning of day t−1; $E_{t-1}X_{t-1}(i)$ is the expectation of $X_{t-1}(i)$ with Y_{t-1} unknown, or as of the beginning of day t−1; and

$$\eta_t(i) = C_{t-1}[Y_{t-1}, X_{t-1}(i)]/V_{t-1}X_{t-1}(i) \equiv E_{t-1}[y_{t-1}x_{t-1}(i)]/E_{t-1}x_{t-1}(i)^2$$

where

$$y_{t-1} = Y_{t-1} - E_{t-1}Y_{t-1}$$

and

$$x_{t-1}(i) = X_{t-1}(i) - E_{t-1}X_{t-1}(i)$$

As is also well known, the residual from the expectation given by equation (17), that is, $Y_{t-1}-E_tY_{t-1}$, is independent of the $X_{t-1}(i)$. We use this result below in deriving the formulae for Π_t.

At the beginning of day t, then, the central bank uses equations like equation (17) to determine the $E_ta_{t-1}(i)$. Its information variables, the counterparts of the $X_{t-1}(i)$, are not the same, however, for all t. For one thing, it does not necessarily use the same variable as its instrument variable on days t and t+1. And an instrument variable cannot be an information variable, for by definition its expected value as of the beginning of day t and its actual value for day t are the same.[9] Also, recall our information assumptions. At the beginning of day t, where t=1, 2, . . . , Q, P_{t-1} and R_{t-1} are known. But M_{t-1} is known only for t=1, ω+1, 2ω+1, . . . , (s−1)ω+1, there being ω days in a week and s weeks in the quarter. And Y_{t-1} is known only for t=1. It is not therefore simply a matter of deriving one set of formulae, the counterparts of equation (17), which hold for all t. There are several sets of formulae and to derive them we have to consider the central bank's situation at the beginning of the first day of the quarter, the second or any subsequent day of the first week and the first day of the second or any other week of the quarter.

A. *Days of the First Week*

The first day of the quarter (t=1) is decidedly special. At its beginning, the central bank—knowing P_0, R_0, M_0, and Y_0—is able actually to calculate a_0. In a manner of speaking, therefore, it uses equations (1), (3), and (4) to determine $E_1a_0=a_0$. And, of course, $\Pi_1=0$.

We now show how the central bank obtains E_ta_{t-1} and Π_t, given that it has $E_{t-1}a_{t-2}$ and Π_{t-1}. At the beginning of day t, where $2\leqq t\leqq\omega$, the central bank gets observations on P_{t-1} and R_{t-1}. One of these observations provides no information about a_{t-1}, though, since the central bank must have used either P or R as its instrument variable for day t−1. If it used P, then by equation (17)

$$(18)\quad E_ta_{t-1}(i) = E_{t-1}a_{t-1}(i) + \phi_t(i)r_{t-1}$$

where

$$r_{t-1} = R_{t-1} - E_{t-1}R_{t-1}$$
$$\phi_t(i) = C_{t-1}[a_{t-1}(i), R_{t-1}]/V_{t-1}R_{t-1}$$

and, by equations (5),

$$(19)\qquad E_{t-1}a_{t-1}(i) = \rho_iE_{t-1}a_{t-2}(i)$$

Now, the reduced-form equation for R_{t-1}, derived from equations (1) and (2), is

$$(20)\quad R_{t-1} = \gamma_0 + (\delta_1/\alpha_1)P_{t-1} + \Gamma'a_{t-1}$$

where

$$\Gamma' = (\beta_1\delta_2, \delta_3/\alpha_1, -\delta_3/\alpha_1)$$

and

$$\gamma_0 = \frac{\beta_0+\alpha_0\beta_1}{1-\alpha_1\beta_1}$$

Therefore, by equation (19),

$$(21)\qquad E_{t-1}R_{t-1} = \gamma_0 + (\delta_1/\alpha_1)P_{t-1} - \Gamma_\rho' E_{t-1}a_{t-2}$$

[9] We show below, however, that E_ta_{t-1} and Π_t are the same, whether P was used as the instrument variable for day t−1 or R was.

where

$$\Gamma_\rho' = (\rho_1\beta_1\delta_2,\ \rho_2\delta_3/\alpha_1,\ -\rho_3\delta_3/\alpha_1)$$

Since r_{t-1} is the difference between R_{t-1} and $E_{t-1}R_{t-1}$, it follows from equations (5) and (7) and (20), and (21) that

$$(22) \qquad r_{t-1} = \Gamma_\rho' e_{t-1} + \Gamma' u_{t-1}$$

Also, by equations (5), (7), and (19),

$$(23) \quad a_{t-1}(i) - E_{t-1}a_{t-1}(i) = \rho_i e_{t-1}(i) + u_{t-1}(i),$$

so

$$(24) \quad C_{t-1}[a_{t-1}(i), R_{t-1}] = \Gamma_i Vu(i) + \rho_i \Gamma_\rho' \Pi_{t-1}(i)$$

where Γ_i is the ith element of Γ and $\Pi_{t-1}(i)$ is the ith column of Π_{t-1},

$$(25) \quad V_{t-1}R_{t-1} = \Gamma'\Phi\Gamma + \Gamma_\rho'\Pi_{t-1}\Gamma_\rho$$

and

$$(26) \quad \phi_t(i) = \frac{\Gamma_i Vu(i) + \rho_i \Gamma_\rho' \Pi_{t-1}(i)}{\Gamma'\Phi\Gamma + \Gamma_\rho'\Pi_{t-1}\Gamma_\rho}$$

Finally, by equations (5), (18), and (23),

$$(27) \quad e_t(i) = \rho_i e_{t-1}(i) - \phi_t(i) r_{t-1} + u_{t-1}(i)$$

And since the $e_t(i)$ are independent of r_{t-1},

$$E_t(e_t e_t') = E_{t-1}(e_t e_t')$$

In consequence,

$$(28) \quad \Pi_t(i,j) = \rho_i\rho_j \Pi_{t-1}(i,j) + E[u(i)u(j)] - \phi_t(i)\phi_t(j) V_{t-1}R_{t-1}$$

or, by equations (25) and (26), we may derive equation (29) shown below.

If P was the instrument variable for day t−1, then the appropriate expectations for deciding open market policy at the beginning of day t are those of equations (18) and (29). These, however, are also the appropriate expectations if R was the instrument variable for day t−1. If it was, then by equation (17)

$$(30) \quad E_t a_{t-1}(i) = E_{t-1}a_{t-1}(i) + \theta_t(i) p_{t-1}$$

But there is a reduced-form equation for P_{t-1}, obtainable from equation (20),

$$(31) \quad P_{t-1} = -(\alpha_1/\delta_1)(\gamma_0 - R_{t-1} + \Gamma' a_{t-1})$$

so

$$(32) \quad p_{t-1} = -(\alpha_1/\delta_1)(\Gamma_\rho' e_{t-1} + \Gamma' u_{t-1})$$

In words, p_{t-1} is simply a multiple of the discrepancy, r_{t-1}, which would have been observed had P been used as the instrument variable (see equation (22)). And since $a_{t-1} - E_{t-1}a_{t-1}$ is independent of the instrument variable choice and value of day t−1 (see equation (23)), it follows that $\theta(i) = -(\delta_1/\alpha_1)\phi(i)$ and, from this equality, that $E_t a_{t-1}$ and Π_t are the same whether P or R was used as the instrument variable for day t−1.[10]

Thus, assuming known unique values of $E_{t-1}a_{t-2}$ and of Π_{t-1}, values assumed to be independent of policy decisions up through day t−2, we have now found unique values of $E_t a_{t-1}$ and of Π_t. These depend on those prior expectations, on certain functions of day t−1 outcomes, and on the parameters of the structure (including the $Vu(i)$), but not on the day t−1 policy decision. For Π_t, this is clear from inspection of equations (29); for $E_t a_{t-1}$, one must inspect equations (18), (19), (22), and (26). Since we showed at the outset of this section that the $E_1 a_0$ and Π_1 are independent of policy

[10] This result is easily checked. One has only to redo our derivations, starting, however, with equations (30) and (31). It is perhaps not surprising that $E_t a_{t-1}$ and Π_t are the same, no matter which variable, P or R, was chosen as instrument variable at the beginning of day t−1. These variables are linearly related (equation (20)), and the disturbance in the relationship is independent of which variable was chosen as the instrument variable.

$$(29) \quad \Pi_t(i,j) = \rho_i\rho_j\Pi_{t-1}(i,j) + E[u(i)u(j)] - \frac{[\Gamma_i Vu(i) + \rho_i\Gamma_\rho'\Pi_{t-1}(i)][\Gamma_i Vu(j) + \rho_j\Gamma_\rho'\Pi_{t-1}(j)]}{\Gamma'\Phi\Gamma + \Gamma_\rho'\Pi_{t-1}\Gamma_\rho}$$

up through day 0, it follows that $E_t a_{t-1}$ and Π_t are independent of the entire sequence of policy decisions up through day $t-1$. That this is true for Π_t can be seen in a slightly different and perhaps more revealing way. Because all of the parameters of the economic structure, including the $Vu(i)$ are known at the beginning of day 1, the central bank, using equations (29), is at that time able to calculate all Π_t.[11]

Strictly speaking, the proof and the formulae we have provided are only for $1 \leqq t \leqq \omega$. It is easy enough, however, to extend the proof for $t > \omega$. Moreover, as will be apparent, the formulae found here apply for second and subsequent days of all the remaining weeks. Only the formulae for the first day of the second and following weeks are missing.

B. *The First Day of the Second Week*

At the beginning of the first day of the second week, day $\omega+1$, the central bank gets observations on the daily deposit stock for the first week. And knowing $M_1, M_2, \ldots, M_\omega$, it revises its expectations of a_{t-1} for $2 \leqq t \leqq \omega$. It puts itself back at the beginning of day 2 and, using observations on M_1 and P_1 or R_1, calculates revised expectations of a_1, denoted by $E_2^* a_1$. Then it puts itself at the beginning of day 3 and, using observations of M_2 and P_2 or R_2, calculates the revised expectation $E_3^* a_2$. And in the end, having calculated $E_t^* a_{t-1}$ for $2 \leqq t \leqq \omega$, it comes to expectations of a_ω, which is in part what it needs to determine open market policy for day $\omega+1$.

If P was the instrument variable for day $t-1$, then for day t the central bank's information variables are R and M. To use equations similar to equation (17) to determine $E_t^* a_{t-1}$, it has to have information variables which are uncorrelated. Obviously, R_{t-1} and M_{t-1} are correlated. But there are transforms,

$$Z_{t-1}(1) = \overset{*}{m}_{t-1} - \alpha_6 \overset{*}{r}_{t-1} \tag{33}$$

and

$$Z_{t-1}(2) = (\delta_3/\alpha_1) Z_{t-1}(1) + \overset{*}{r}_{t-1} \tag{34}$$

which are uncorrelated.

By equations (4) and (20), the reduced-form equation for M_{t-1} is

$$\begin{aligned} M_{t-1} = \alpha_5 + (\alpha_7 + \alpha_6\delta_1/\alpha_1)P_{t-1} \\ + \alpha_6 \Gamma' a_{t-1} + a_{t-1}(3); \end{aligned} \tag{35}$$

therefore,

$$\begin{aligned} \overset{*}{m}_{t-1} &\equiv M_{t-1} - \overset{*}{E}_{t-1} M_{t-1} \\ &= \alpha_6 \Gamma'(a_{t-1} - \overset{*}{E}_{t-1} a_{t-1}) \\ &\quad + a_{t-1}(3) - \overset{*}{E}_{t-1} a_{t-1}(3) \end{aligned} \tag{36}$$

But since

$$\begin{aligned} a_{t-1}(i) - \overset{*}{E}_{t-1} a_{t-1}(i) \\ = \rho_i \overset{*}{e}_{t-1}(i) + u_{t-1}(i) \end{aligned} \tag{37}$$

it follows that

$$\begin{aligned} \overset{*}{m}_{t-1} = \alpha_6(\Gamma_\rho' \overset{*}{e}_{t-1} + \Gamma' u_{t-1}) \\ + \rho_3 \overset{*}{e}_{t-1}(3) + u_{t-1}(3) \end{aligned} \tag{38}$$

And by the analogue of a previous argument (see the derivation of equation (22)),

$$\overset{*}{r}_{t-1} = \Gamma_\rho' \overset{*}{e}_{t-1} + \Gamma' u_{t-1}, \tag{39}$$

so

$$Z_{t-1}(1) = \rho_3 \overset{*}{e}_{t-1}(3) + u_{t-1}(3) \tag{40}$$

and

$$Z_{t-1}(2) = \Gamma_\rho^{*\prime} \overset{*}{e}_{t-1} + \Gamma^{*\prime} u_{t-1} \tag{41}$$

where

$$\Gamma^{*\prime} = (\beta_1\delta_2, \delta_3/\alpha_1, 0)$$

and[12]

$$\Gamma_\rho^{*\prime} = (\rho_1\beta_1\delta_2, \rho_2\delta_3/\alpha_1, 0)$$

[11] Note that there is no circularity of reasoning, since in this section we have not assumed optimality of policy, nor, in fact, anything about policy, except that each day either P or R is used as the instrument variable.

[12] That $Z_{t-1}(1)$ and $Z_{t-1}(2)$ are uncorrelated is estab-

Now, since $Z_{t-1}(1)$ and $Z_{t-1}(2)$ are uncorrelated,

$$E_t^* a_{t-1}(i) = E_{t-1}^* a_{t-1}(i) + \sum_{k=1}^{2} \phi_t^*(i, k) Z_{t-1}(k) \tag{42}$$

where

$$\phi_t^*(i, k) = \frac{C_{t-1}^*[a_{t-1}(i), Z_{t-1}(k)]}{V_{t-1}Z_{t-1}(k)}$$

As is readily verified,

$$\phi_t^*(i, 1) = \begin{cases} 0 & \text{for } i = 1, 2 \\ 1 & \text{for } i = 3 \end{cases}$$

and

$$\phi_t^*(i, 2) = \begin{cases} \dfrac{\rho_i \Gamma_p^{*\prime} \Pi_{t-1}^*(i) + \Gamma_i^{*\prime} V u(i)}{\Gamma_p^{*\prime} \Pi_{t-1}^* \Gamma_p^* + \Gamma^{*\prime} \Phi \Gamma^*} & \text{for } i = 1, 2 \\ 0 \quad \text{for } i = 3 \end{cases}$$

where $\Pi_{t-1}^*(i)$ is the ith column of Π_{t-1}^*, the variance-covariance matrix of e_{t-1}^* and Γ_i^* is the ith element of Γ^*.

And by equations (37) and (42),

$$e_t^*(i) = \rho_i e_{t-1}^*(i) - \sum_{k=1}^{2} \phi_t^*(i, k) Z_{t-1}(k) + u_{t-1}(i), \tag{43}$$

so

$$\Pi_t^*(i, j) = \begin{cases} \rho_i \rho_j \Pi_{t-1}^*(i, j) + E[u(i)u(j)] \\ \quad - \phi_t^*(i, 2)\phi_t^*(j, 2) V_{t-1}^* Z_{t-1}(2) & \text{for } i, j \neq 3 \\ 0 \quad \text{for } i = 3 \text{ or } j = 3 \end{cases} \tag{44}$$

There is the possibility, however, of the central bank having used R as its instrument variable for day t−1. If it did, then P and M are its information variables. The uncorrelated transforms are

$$Z_{t-1}(3) = m_{t-1}^* + (\alpha_6 \delta_1/\alpha_1) p_{t-1}^* \tag{45}$$

and

$$Z_{t-1}(4) = (\delta_3/\delta_1) Z_{t-1}(3) - p_{t-1}^* \tag{46}$$

where p_{t-1}^* is the analogue of p_{t-1}. But $Z_{t-1}(3) = Z_{t-1}(1)$ and $Z_{t-1}(4) = (\alpha_1/\delta_1) Z_{t-1}(2)$, so $E_t^* a_{t-1}$ is the same, whether P was chosen as the instrument variable at the beginning of day t−1 or R was.

The central bank, at the beginning of day $\omega+1$, uses equations (42) and (44)—but, as already indicated, repeatedly—to get the necessary expectations. Its situation at the beginning of day $2\omega+1$, or any other first day of the week, is exactly the same, however, as its situation at the beginning of day $\omega+1$, so we have now completed our task of providing all the required expectations formulae.

C. *A Generalization*

In this subsection, we provide the expectations formulae for the central bank which is constrained by an economic structure which, although also linear, is more complicated than that of Section I. These formulae are easy generalizations of those derived above. Setting them out, we nevertheless make an important point: that in deciding open market policy, the optimizing central bank "looks at everything." This point has been made by others, but perhaps not fully justified.

We assume here that there are K inter-

---

lished by the following argument. Because a_0 is known, $e_1^*(3) = 0$. Therefore, by equation (40), $Z_1(1) = u_1(3)$. And since $Z_1(1)$ is known at the beginning of day 2, $u_1(3)$ is too. Consequently,

$$E_2^* a_1(3) = E_2^*[\rho_3 a_0(3) + u_1(3)] = a_1(3)$$

and $e_2^*(3) = 0$, so $u_2(3)$ is known as of the beginning of day 3. More generally, $Z_t(1) = u_t(3)$ and $e_t^*(3) = 0$ for $t \leq \omega$, so as alleged, $Z_{t-1}(1)$ and $Z_{t-1}(2)$ are uncorrelated. It is a consequence of the particular economic structure of Section I, however, that $e_t^*(3) = 0$ and that $u_t(3)$ is known as of the beginning of day t+1. In general, it is not possible to identify any of the day t disturbances at the beginning of day t+1. But this does not matter, since for any linear economic structure there exist uncorrelated transforms of all the natural information variables that contain all the information.

est rates, the kth of which is $R_t(k)$, and N monetary aggregates, the nth of which is $M_t(n)$. At the beginning of day t where $t=1, 2, \ldots, Q$ the central bank observes P_{t-1} and $R_{t-1}(k)$, $k=1, 2, \ldots, K$. And at the beginning of day t where $t=1, \omega+1, \ldots, (s-1)\omega+1$ it observes $M_{t-1}(n)$, $M_{t-2}(n), \ldots, M_{t-\omega}(n)$, $n=1, 2, \ldots, N$.

It is enough to give the formulae which apply if P was the instrument variable for day $t-1$. Thus,

$$E_t a_{t-1}(i) = E_{t-1}a_{t-1}(i) + \sum_{k=1}^{K} \psi_t(i, k) Z_{t-1}(k) \tag{47}$$

for $i=1, 2, \ldots, I \geq K+N+1$ and $t=2, 3, \ldots, \omega, \omega+2, \ldots, Q$ where $Z_{t-1}(k)$, $k=1, 2, \ldots, K$ are independent linear transforms of the $r_{t-1}(k)$ and

$$\psi_t(i, k) = \frac{C_{t-1}[a_{t-1}(i), Z_{t-1}(k)]}{V_{t-1}Z_{t-1}(k)}$$

Also,

$$\Pi_t(i, j) = \rho_i\rho_j\Pi_{t-1}(i, j) + E[u(i)u(j)] - \sum_{k=1}^{K} \psi_t(i, k)\psi_t(j, k) V_{t-1}Z_{t-1}(k) \tag{48}$$

And, finally, for $t=\omega+1, 2\omega+1, \ldots, (s-1)\omega+1$,

$$E_t a_{t-1}(i) = \overset{*}{E}_{t-1}a_{t-1}(i) + \sum_{k=1}^{K+N} \overset{*}{\psi}_t(i, k) \overset{*}{Z}_{t-1}(k) \tag{49}$$

where the $Z^*_{t-1}(k)$ are independent linear transforms of the $r^*_{t-1}(k)$ and the $m^*_{t-1}(k)$ and

$$\overset{*}{\psi}_t(i, k) = \frac{\overset{*}{C}_{t-1}[a_{t-1}(i), \overset{*}{Z}_{t-1}(k)]}{V^*_{t-1}Z^*_{t-1}(k)}$$

and

$$\overset{*}{\Pi}_t(i, j) = \rho_i\rho_j\overset{*}{\Pi}_{t-1}(i, j) + E[u(i)u(j)] - \sum_{k=1}^{K+N} \overset{*}{\psi}_t(i, k)\overset{*}{\psi}_t(j, k) \overset{*}{V}_{t-1}\overset{*}{Z}_{t-1}(k) \tag{50}$$

In deciding open market policy at the beginning of day t, the optimizing central bank does then use all of its most recent observations, those on interest rates and, if there are any, those on monetary aggregates. So much is clear from equations (47) and (49). Nor does it respond to or "weight" its most recent observations in an arbitrary way. The weights are determined by the economic structure.

III. An Alternative Strategy

We have but one remaining task: to evaluate an alternative strategy for open market operations. This strategy, which we think of as the intermediate target variable strategy, is in our judgment the strategy the *FOMC* would be using if it were not at all concerned about day-to-day and week-to-week changes in interest rates and if the simplifying assumptions of Section I were satisfied. It is, we believe, essentially like the *FOMC*'s actual strategy.[13] That of course is why we bother to evaluate it and why in this section we portray the *FOMC* as using it.

The *FOMC* has a target path for Y—a sequence of target values, that is, one for each day of the quarter—which is determined at the very beginning of the quarter. As above, we take this target path to be $Y(t)=\bar{Y}$, for all t. The question is how the *FOMC* goes about trying to realize its objective, which is equality between actual Y and $\bar{Y}$ for every day t of the quarter. To facilitate understanding, we first provide a

[13] In formulating our alternative or intermediate target variable strategy, we tried to approximate the *FOMC* strategy of 1970–71, which has been described by, among others, Stephen Axilrod, Alan Holmes, and Paul Meek and Rudolf Thumberg. There have been indications that early in 1972 the *FOMC* changed its strategy, but not so far as we are concerned in an essential way. What the *FOMC* evidently did in January–February 1972 was switch from using the federal funds rate as its operating variable to using a reserves aggregate. To all appearances, however, it did not give up the use of certain variables (for example, the narrowly defined stock of money) as intermediate target variables. And here that is what is important.

rather loose description of how the *FOMC* proceeds. Thereafter, though, we provide a precise description. We do this by actually deriving the sequence of desired or target values for the *FOMC*'s instrument variable, denoted by $\bar{P}(t)$, which is implied by $\bar{Y}$. Then we compare $\bar{P}(t)$ and $\tilde{P}(t)$, the latter being our optimal value of P for day t, and in so doing evaluate the *FOMC* strategy.

The *FOMC* first decides on a target value for Y. Then, using this value, it goes on and determines a target path or sequence of target values for M, its intermediate target variable. And, what is most important, it does this at the very beginning of the quarter. Nor does this sequence of target values ever get revised. Whatever happens during the quarter, the *FOMC* continuously strives for equality between actual M and successive values of the predetermined target sequence. So what would seem to be its intermediate objective—to repeat, achieving equality between actual and target values of M—is in a very real sense its ultimate objective. But it goes about trying to achieve its intermediate objective in a particular way. At the beginning of each week, it determines a sequence of target values for R, its operating variable, a sequence which, like the quarterly sequence of M, never gets revised. Over the week, the *FOMC* tries, by managing its portfolio, to achieve equality between actual R and the successive values of the predetermined sequence. It thus has not only an intermediate objective but also what we might call a proximate (weekly) objective.

Unless the economic structure structure is very special, however, it is not optimal to proceed in this way, using an intermediate target variable (or an operating variable). To put the point differently, it is not optimal to use predetermined sequences of target values, sequences which, although new observations are received, never get revised. That is what we now show. We first derive what for the intermediate target variable or *FOMC* strategy is the optimal sequence of portfolio values.

The target path of M is the day 1 path of expected values, E_1M_t, $t=1, 2, \ldots, Q$, determined by the equalities $E_1Y_t=\bar{Y}$ and equations (1), (3), and (5). If actual M were always on this path and if there were no disturbances, then actual Y would always be on its target path. The *FOMC*'s intermediate objective is therefore that

$$E_TM_{T+h} = E_1M_{T+h} \tag{51}$$

be satisfied for all $T=1, \omega+1, 2\omega+1, \ldots, (s-1)\omega+1$ and $h=0, 1, \ldots, \omega-1$. The term E_TM_{T+h} is of course the expected value of M_{T+h} as of the beginning of day T, the first day of some week.

But as we have already indicated, the *FOMC* goes about ensuring that equations (51) are satisfied in a particular way. At the beginning of day T, it determines a sequence of target values for R, its operating variable, a sequence denoted by $\bar{E}_TR_{T+h}$, where $h=0, 1, \ldots, \omega-1$, which satisfies equations (51). This sequence is obtained as follows. By equations (1), (3), and (5)

$$M_t = \epsilon_0 + \epsilon_1R_t + A_\rho' a_{t-1} + A'u_{t-1} \tag{52}$$

where

$$\begin{aligned}\epsilon_0 &= \alpha_2 + \alpha_0\alpha_3\\ \epsilon_1 &= \alpha_4 + \alpha_1\alpha_3\\ A_\rho' &= (\rho_1\alpha_3, \rho_2, 0)\end{aligned}$$

and

$$A' = (\alpha_3, 1, 0)$$

Therefore,

$$E_tM_t = \epsilon_0 + \epsilon_1E_tR_t + A_\rho' E_ta_{t-1} \tag{53}$$

and

$$\begin{aligned}&E_TM_{T+h} - E_1M_{T+h}\\ &\quad= \epsilon_1(E_TR_{T+h} - E_1R_{T+h})\\ &\qquad+ A_\rho'(E_Ta_{T+h-1} - E_1a_{T+h-1})\end{aligned} \tag{54}$$

Equations (54), together with equations (51), give the sequence, so

$$\bar{E}_T R_{T+h} = E_1 R_{T+h} - \frac{A_\rho'}{\epsilon_1}(E_T a_{T+h-1} - E_1 a_{T+h-1}) \tag{55}$$

And since $E_1 R_{T+h}$ is the expected value of R_t obtained from equation (6) with $E_1 Y_{T+h}$ set equal to $\tilde{Y}$,

$$\bar{E}_T R_{T+h} = [\tilde{Y} - \alpha_0 - \rho_1 E_1 a_{T+h-1}(1) - \frac{A_\rho'}{\epsilon_1}(E_T a_{T+h-1} - E_1 a_{T+h-1})]/\alpha_1 \tag{56}$$

The *FOMC*'s proximate objective is that

$$E_{T+h} R_{T+h} = \bar{E}_T R_{T+h} \tag{57}$$

be satisfied for all $h=0, 1, \ldots, \omega-1$. The expected value of R_{T+h} as of the beginning of day $T+h$ is to equal the target value for R_{T+h}, which is the expected value as of the beginning of day T. Why this proximate objective will be clear. Along the path $\bar{E}_T R_{T+h}$, the expected value of M_{T+h} as of the beginning of day T is equal to the target value of M_{T+h}, the expected value as of the beginning of day 1.

And now to get the sequence of values for the *FOMC*'s portfolio, denoted by $\bar{P}_{T+h}$, where $h=0, 1, \ldots, \omega-1$, which is open market policy for the week. It is the sequence which satisfies equations (57). But by equations (20),

$$E_{T+h} R_{T+h} = \gamma_0 + \frac{\delta_1}{\alpha_1} P_{T+h} + \Gamma_\rho' E_{T+h} a_{T+h-1} \tag{58}$$

So, by equations (57) and equations (56), we derive an expression for $\bar{P}_{T+h}$ in equation (59) shown below.

We are finally in a position to evaluate the intermediate target variable strategy. This we do my comparing $\bar{P}_{T+h}$ and $\tilde{P}_{T+h}$, the value of the *FOMC* portfolio which minimizes expected loss. Since

$$\Delta_\rho' - \alpha_1 \Gamma_\rho' = (\rho_1, 0, 0)$$

It follows from equations (15) and (59) that

$$\bar{P}_{T+h} - \tilde{P}_{T+h} = \frac{\rho_1}{\delta_1}(E_{T+h} a_{T+h-1} - E_1 a_{T+h-1}) - \frac{\alpha_1}{\delta_1 \epsilon_1} A_\rho' (E_T a_{T+h-1} - E_1 a_{T+h-1}) \tag{60}$$

But by equation (60), $\bar{P}_{T+h} - \tilde{P}_{T+h}$ vanishes for $T=1, \omega+1, \ldots, (s-1)\omega+1$ and $h=0, 1, \ldots, \omega-1$ if, and only if, $\rho_1=\rho_2=0$. Thus, except on the first day of the quarter, the *FOMC*'s way of determining its portfolio (in open market policy) is optimal if, and only if, $\rho_1=\rho_2=0$. Yet, $\rho_1=\rho_2=0$ means that the disturbances of equations (1) and (3) are not serially correlated or, more generally, that the past influences neither the private sector's desired expenditure for newly produced goods and services nor its desired stock of demand deposits. And this, it would seem, is not what most economists believe.

There are economists (we might call them monetarists) who have, though, advocated that the *FOMC* use M as, in effect, its ultimate target variable or, what comes to the same thing, that it start off the quarter by determining a target path, $E_1 M_t$, and strive through the quarter to stay on this path. It would therefore seem worthwhile to get conditions under which it is optimal to use M rather than Y as the target variable. We do so, however, on the

$$\bar{P}_{T+h} = \frac{1}{\delta_1}(\tilde{Y} - \delta_0) - \frac{\rho_1}{\delta_1} E_1 a_{T+h-1} - \frac{\alpha_1}{\delta_1}\Gamma_\rho' E_{T+h} a_{T+h-1} - \frac{\alpha_1}{\delta_1 \epsilon_1} A_\rho' (E_T a_{T+h-1} - E_1 a_{T+h-1}) \tag{59}$$

assumption that the value of P_t is determined directly from

$$(61) \qquad E_t M_t = E_1 M_t$$

The value so determined is $\hat{P}_t$. And, by equation (35),[14]

$$(62) \qquad \hat{P}_t - \tilde{P}_t = \left(\frac{1}{\delta_1}\Delta_\rho' - \frac{\alpha_1}{\delta_1}\Gamma_\rho' - \frac{\alpha_1}{\delta_1\epsilon_1}A_\rho'\right) \cdot (E_t a_{t-1} - E_1 a_{t-1})$$

But since

$$\left(\frac{1}{\delta_1}\Delta_\rho' - \frac{\alpha_1}{\delta_1}\Gamma_\rho' - \frac{\alpha_1}{\delta_1\epsilon_1}A_\rho'\right) = \left[\rho_1\left(1 - \frac{\alpha_1\alpha_3}{\alpha_4 + \alpha_1\alpha_3}\right), -\rho_2\frac{\alpha_1}{\alpha_4 + \alpha_1\alpha_3}, 0\right]$$

It is immediate that $\hat{P}_t = \tilde{P}_t$ if, and only if, $\alpha_4, \rho_2 = 0$ or $\rho_1, \rho_2 = 0$.

Those who would have the *FOMC* use M as in effect its ultimate target variable evidently believe therefore either that the past influences neither the private sector's desired expenditures nor its desired stock of deposits or, what is more likely, that neither the past nor the rate of interest influences the private sector's desired stock of deposits. This does not seem terribly likely. But even if we accepted that the private sector's desired stock of deposits is not influenced by the past or by the rate of interest, we would not get optimality for the *FOMC* strategy. It involves the substitution of M for Y as the ultimate target variable. It also involves choosing a sequence of target values for its operating variable, R, at the beginning of each week. And that is what is not optimal, even on the assumption that the private sector's desired stock of deposits is not influenced by the past or by the rate of interest. The proof of this is easy. By equation (60), if $\alpha_4, \rho_2 = 0$ and $\rho_1 \neq 0$, then $\bar{P}_t \neq \tilde{P}_t$.

We must point out before concluding, that in evaluating the intermediate target variable or *FOMC* strategy, we have been concerned entirely with the implied value of the given instrument variable P. But $\bar{P} = \tilde{P}$ is only a necessary condition for optimality. The implied value for P could be $\bar{P}$ and still the *FOMC* strategy might not be optimal. For there is another possible instrument variable, R, which the *FOMC* could be well advised to use. Some have denied this possibility. By inequalities (16), however, $\alpha_4, \rho_2 = 0$ (or $\rho_1 = \rho_2 = 0$) does not imply that P is the better instrument variable.

IV. Conclusion

We began by setting out certain simplifying assumptions and we must now, in ending, be clear on what it means, our having made these assumptions. So far as we know, what we have provided in this paper is not the optimal strategy for open market operations but only an optimal strategy. Had we started out with different assumptions, we likely would not have come to exactly the strategy of Sections I and II. And we might not have found the intermediate target variable strategy of Section III to be generally sub-optimal. We are disposed to think, though, that we would have.[15]

Some readers may already have dismissed our conclusion about the sub-optimality of the intermediate target variable strategy (or, more particularly, the *FOMC* strategy) as being of no practical significance. In deriving it, did we not use what even we described as an unrealistic assumption—to wit, that the central

[14] It will be noted that if $h=0$, then equations (60) and (62) are the same. What this means is that, given M as the ultimate target variable, the *FOMC* strategy is optimal for the first day of the week.

[15] Unless of course we had started out by assuming serially uncorrelated residuals. (Recall our having shown that with serially uncorrelated residuals the strategy of Section III is optimal.) We have in mind the possibility of having started out by allowing for uncertainty about the values of the constants and coefficients of the economic structure.

bank knows with certainty that version of the economic structure which holds over the decision period or, as we would say, over days? Think of the *FOMC*. Its decision period is the calendar day. And can it seriously be maintained that the *FOMC* knows with certainty that version of the *U.S.* economic structure which determines, *inter alia*, daily *GNP* and daily average interest rates? For one thing, the U.S. Department of Commerce has yet to provide estimates of monthly *GNP*, to say nothing of daily *GNP*.

Yet, even if we had been confident of our ability to manage with any assumptions whatever, we would not have started off by assuming that the central bank knows nothing about that version of the economic structure which holds over the decision period, but rather that it is uncertain about constant and coefficient values. And why? Well, consider the *FOMC*'s situation. It presumably knows, although not with certainty, that version of the *U.S.* economic structure which holds over calendar quarters or which determines calendar-quarter *GNP* and, say, the quarterly average stock of money. All the various versions of any economic structure must, however, be consistent. So there is a version of the *U.S.* economic structure which holds over the calendar day and which, being implied by the version holding over the calendar quarter, the *FOMC* could come to know, although again not with certainty.[16]

We are not saying that the *FOMC* actually knows that version of the *U.S.* economic structure which determines calendar-day *GNP*, but only that by spending sufficient time and money it could. This being so, it would not have been unreasonable if we had started out by assuming that the central bank knows, if with uncertainty, the decision period version of the economic structure. At issue therefore is whether on this assumption the intermediate target variable strategy of Section III is optimal. And as we have already indicated, we are doubtful. The point is that the use of an intermediate target variable necessarily involves some waste of information which, if initial conditions matter, is valuable. And it is difficult to imagine the waste of valuable information being optimal, even for a central bank which is uncertain about the appropriate version of the economic structure constraining it.[17] We must admit, however, to having no idea how far from optimal the *FOMC*'s strategy is. Nor will we until we have determined the optimal open market strategy for a central bank which is uncertain about the version of the economic structure which holds over its decision period

APPENDIX

Deposit Stock Observation Errors

One of our information assumptions was that at the beginning of the first day of each week, the central bank gets error-free observations on the deposit stock for all the days of the week just past. We might, however, have made a more realistic assumption: that it gets only estimates for the days of the week just past and, at the same time, error-free observations for the days of the week before that. Actually, it would not have made all that much difference if we had; our optimal strategy would not have been much different. But since the *FOMC* initially gets estimates of the daily deposit stock, it is per-

[16] Quite obviously, there may be several different consistent versions of the structure which holds over the decision period. There is one version, however, of which all of these versions are special cases.

[17] We note in passing, for the benefit of those who might want to insist (in our view, quite mistakenly) on complete ignorance of the decision period version of the economic structure, that this can hardly imply use of one or several intermediate target variables. Such use necessarily involves changing the value of the instrument variable from decision period to decision period. But if ignorance is complete, there is no way, other than by guessing, of determining the appropriate new value for the instrument variable.

haps worth explaining how the introduction of observation errors changes the optimal open market strategy.

To anticipate, the central bank's expectations change with the introduction of deposit stock observation errors. It gets its expected values of, say, initial conditions exactly as it would if there were no observation errors, by using observations on all information variables and such estimates of the deposit stock as are available. But, again, its expectations are not the same as they would be if there were no errors of observation. Nor is the sequence of optimal values for the instrument variable(s); as will be recalled, expectations of initial conditions partly determine optimal policy. The central bank proceeds, however, exactly as it would if there were no observation errors; in determining its optimal instrument variable and the optimal value thereof, it uses formulae entirely analogous to those given in the text.

The estimate of the deposit stock for day t is

$$\hat{M}_t = M_t + v_t$$

where v_t is normally distributed,

$$E_t v_t = 0$$

$$E_t(v_t v_{t+h}) = \begin{cases} V_t v_t = Vv & \text{for } h = 0 \\ 0 & \text{otherwise} \end{cases}$$

and

$$E_t[v_t u_t(i)] = 0 \quad \text{for all } i$$

The central bank starts off the first day of the quarter with full information, except about the daily deposit stock. For the last week of the quarter just past, it has only estimates of the daily deposit stock. Not knowing M_0, it therefore goes back, as it were to the beginning of the day $-(\omega-1)$, the first day of the last week of the quarter just past, and calculates $a_{-\omega}$. (Since it has full information, determining these initial conditions is a simple exercise in arithmetic.) Then, using the calculated $a_{-\omega}$, observations of $Y_{-\omega+1}$ and $P_{-\omega+1}$ or $R_{-\omega+1}$ and the estimate $\hat{M}_{-\omega+1}$, it calculates the revised expectations $E^*_{-\omega+2}a_{-\omega+1}$ and $\hat{\Pi}_{-\omega+2}$. There is a generalization of equation (17) which does not require that the X's be uncorrelated and a particular version of this generalization is what the central bank uses. It proceeds day by day through the last week of the quarter just past and ends with the needed expectations $E_1^* a_0$ and $\hat{\Pi}_1$ where as can be shown

$$\hat{\Pi}_1 = \tau_1 Vv \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}, \quad 0 < \tau_1 < 1$$

Having come up with $E_1^* a_0$ and $\hat{\Pi}_1$, the central bank is able to determine open market policy for the first day of the current quarter. Using $\hat{\Pi}_1$, it determines which is the better instrument variable, P or R.[18] And using $E_1^* a_0$, it determines the optimal value of its chosen instrument variable.

At the beginning of day t, where t=2, 3, . . . , ω, the central bank uses equations which are just like equations (18) and (29), except that $\hat{\Pi}_t$ appears in place of Π_t.

At the beginning of the first day of the second week, the central bank observes the daily deposit stock for the days of the last week of the previous quarter and gets estimates for the days of the first week of the current quarter. It therefore starts off the second week knowing a_0. So it is able, using the known a_0 and the estimates $\hat{M}_1, \hat{M}_2, \ldots, \hat{M}_\omega$, to revise its expectations of a_t, t=1, 2, . . . , ω and thus come to an expected value for $a_{\omega+1}$.

Because $\hat{M}_t$ is only an estimate of M_t,

$$\hat{m}_t^* = \hat{M}_t - E_t^* \hat{M}_t = m_t^* + v_t$$

It follows that

$$\hat{z}_t(1) \equiv \hat{m}_t^* - \alpha_6 v_t^* = Z_t(1) + v_t$$

and

$$\hat{z}_t(2) \equiv (\delta_3/\alpha_1)\hat{z}_t(1) + r_t^* = Z_t(2) + (\delta_3/\alpha_1)v_t$$

But then $\hat{z}_t(1)$ and $\hat{z}_t(2)$ are correlated, so with observation errors the revised expecta-

[18] It might be expected that by increasing Vv, R can be made the optimal instrument variable for day 1. And, indeed, if $\rho_2 \neq \rho_3$, that is so. If $\rho_2 = \rho_3$, however, then inequalities (16) apply unchanged.

tions of the $a_{t-1}(i)$, or of initial conditions for day t, are[19]

$$E_t^* a_{t-1}(i) = E_{t-1}^* a_{t-1}(i) + \sum_{k=1}^{2} \hat{\phi}_t(i,k)\hat{z}_{t-1}(k)$$

where

$$\begin{bmatrix} \hat{\phi}_t^*(i,1) \\ \hat{\phi}_t^*(i,2) \end{bmatrix} = \left[E_{t-1}^*(\hat{z}_{t-1}\hat{z}_{t-1}')\right]^{-1} C_{t-1}^*\left[\hat{z}_{t-1}, a_{t-1}(i)\right]$$

and

$$\hat{z}_{t-1}' = \left[\hat{z}_{t-1}(1), \hat{z}_{t-1}(2)\right]$$

Finally,[20]

$$\hat{\Pi}_t^*(i,j) = \rho_i\rho_j \Pi_{t-1}^*(i,j) + E[u(i)u(j)] - \hat{\phi}_t^*(i)' E_{t-1}^*(\hat{z}_{t-1}\hat{z}_{t-1}')\hat{\phi}_t^*(j)$$

[19] These equations are substitutes for equations (42).

[20] These equations substitute for equations (44). Note that $\hat{\Pi}_1^* = 0$.

REFERENCES

S. Axilrod, "The FOMC Directive as Structured in the Late 1960s: Theory and Appraisal," in *Open Market Policies and Operating Procedures: Staff Studies*, Washington 1971.

A. Holmes, "A Day at the Trading Desk," *Monthly Rev.*, Fed. Reserve Bank New York, Oct. 1970, *52*, 234–38.

J. H. Kareken, "The Optimum Monetary Instrument Variable," *J. Money, Credit, Banking* Aug. 1970, *2*, 385–90.

P. Meek and R. Thumberg, "Monetary Aggregates and Federal Reserve Open Market Operations," *Monthly Rev.*, Fed. Reserve Bank New York, Apr. 1971, *53*, 80–89.

W. Poole, "Optimal Choice of Monetary Policy Instruments in a Simple Stochastic Macro Model," *Quarterly J. Econ.*, May 1970, *84*, 197–216.

A Dynamic Theory of Comparative Advantage

By ROGER W. KLEIN*

The purpose of this paper is to explain two aspects of international trade in the post-World War II period. First, *U.S.* firms, operating within the United States, are often the first to innovate, develop, and produce new products. Second, the *U.S.* advantage in producing a new product often eventually shifts to another country, implying that the product is produced by either a foreign firm or a foreign subsidiary of a *U.S.* firm.[1] By explaining why and when *U.S.* firms establish foreign subsidiaries, this paper therefore attempts to explain why multi-national *U.S.* firms are formed.

A dynamic theory of comparative advantage, which helps explain the multinational firm, is formulated in Section I. It differs from previous formulations by recognizing that firms engage in Research and Development (*R&D*) to learn about a product's production process. By incorporating this learning activity into the traditional production function and specifying its dynamic properties, a dynamic theory of comparative advantage will be developed. Section II argues that this theory applies to the United States, and Section III derives a test of this theory. In Section IV, in addition to discussing the results of this test, the theory developed here will be compared with alternative explanations of the behavior described above.

I. A Dynamic Theory of Comparative Advantage

The Importance of Learning

One can distinguish two sources of a country's comparative advantage in producing a new product. First, a country might derive its advantage from employing learning factors (scientists and engineers) to learn about and improve the product's production process.[2] Second, given that the production process is completely known, a country might have an advantage in employing capital and labor to produce it. For reasons which will become apparent below, denote the former as the learning advantage and the latter as the static advantage. Then, in a two-country world, assume country I has a learning advantage while country II has a static advantage in producing commodity *a*. If the learning advantage dominates the static advantage, country I will have a comparative advantage in *a*. However, if country I's learning advantage declines over time, as either opportunities for additional learning are exhausted or as foreign countries learn about *a*'s production process, comparative advantage in *a* may shift to country II. This argument, therefore, suggests that one can formulate a dynamic theory of comparative advantage by formalizing the learning activity.

* Assistant professor of economics, Purdue University. This paper was adapted from my Ph.D. thesis, written at Yale under Tibor Scitovsky, David Grether, and John McGowan. I am very grateful for their helpful comments and suggestions. I am also very much indebted to Henry Grabowsky for supplying data on *R&D* expenditures of pharmaceutical firms and to Paul de Haen for supplying data on new products. Burton Klein and John Carlson also provided helpful comments.

[1] See the Department of Commerce.

[2] Since many aspects of the production process for a new product are unknown, firms allocate resources (scientists and engineers) to learn about the product's production process.

Assumptions

Assume that all firms are perfectly competitive and that firm i has the following production function for a specific commodity:

$$q_i = S_i(Z) \cdot G_i(K, L)$$
$$Z \equiv (z_1 \ldots, z_n)$$

S_i is firm i's knowledge function, which depends on a vector of cumulative variables. When a distinction is important, i will be denoted IN for the innovator (assumed to be a domestic firm), d for domestic firms, and f for foreign firms. Then assume:

(a) $\partial S_i/\partial z_j > 0; \quad \partial^2/S_i/\partial z_j^2 < 0$

The marginal products of learning factors are positive and diminishing.

(b) $\text{Maximum}\,(S_i) = \underset{Z\to\infty}{\text{Limit}}\,(S_i) = 1$

There is an upper, asymptotic bound on the knowledge function for a specific product.[3]

(c) $S_d(\overline{Z}) > S_f(\overline{Z}) > 0$

For a fixed level of learning factors, $\overline{Z}$, all domestic firms have a learning advantage over foreign firms.

(d) Domestic firms must produce in the domestic country to realize the advantage stated in (c).

(e) The following notation depicts long-run cost functions,[4] exclusive of $R\&D$ expenditures, at time t for the ith firm that entered the market at time s:

AC^i_{st} = Average Cost

$NLAC^i_{st}$ = Average Cost when learning ceases, i.e., when $S_i \approx 1$

Assume:

i) min $(NLAC^f_{st})$ < min $(NLAC^d_{st})$, for all s, t. Foreign firms (f) have an advantage in terms of K and L (the static advantage). For simplicity, assume these average cost functions are constant.

ii) In nominal terms, the domestic learning advantage (c) initially is sufficiently larger than the foreign static advantage to imply:

$$\min\,(AC^d_{st}) < \min\,(AC^f_{st}),$$
$$\text{for s} \leq \text{t}, \quad \text{and} \quad \text{t} < \text{v}$$

While the learning advantage dominates the static advantage (until time v), domestic firms can produce the commodity at a lower minimum average cost than foreign firms.

(f) For simplicity, assume:

$$S_i(z_1, \ldots, z_m) = S_{i1}(z_1) \cdot S_{i2})\,(z_2) \ldots S_{im}(z_m)$$

(g) All countries have identical preference maps and either identical incomes or homothetic preference maps. To emphasize the dynamics of learning, this assumption is made to guarantee that demand factors do not determine comparative advantage.

The second portion of the production function, $G_i(K, L)$, will be termed the static production function; it is the production function when the stock of knowledge ceases to increase ($S_i \approx 1$) and depends on capital, K, and labor, L. In addition to the properties usually attributed to production functions, assume G exhibits constant returns to scale and is the same for all countries.

Formulation of the Theory

Let a particular firm have the following production function at time t for a given product, a:

[3] When $S \approx 1$, $G_i(K, L)$ becomes the production function. If K and L were immobile among countries, comparative advantage would now be determined by Heckscher-Ohlin conditions. Indeed, one could describe a Heckscher-Ohlin world as one in which all factors are immobile and the stock of knowledge is constant.

[4] These are long-run cost functions in the sense that captial, K, and production labor, L_p, are completely variable. Management, L_m, is assumed to be fixed over the firm's long-run decision horizon. These cost curves will then be U-shaped.

$$q_t = S(Q_{t-1}, RD_{t-1}) \cdot G(K_t, L_t) \text{ where:}$$

$$Q_{t-1} \equiv \text{cumulative output} \equiv \sum_{k=1}^{t-1} q_k$$

$$RD_{t-1} \equiv \text{cumulative, real } R\&D \equiv \sum_{k=1}^{t-1} rd_k$$

Q_{t-1} reflects "learning by doing" and is assumed to be exogenous (see Kenneth Arrow, pp. 165–73), while RD_{t-1} reflects *R&D*-learning (allocating resources to learn about a product's production process) and is assumed to be endogenous.

Since rd_t is an endogenous variable affecting all future production levels, a profit maximizing firm chooses rd_t, q_t, K_t, and L_t to maximize its discounted profit stream. This behavior can be simplified because these values of q_t, K_t, and L_t also maximize period t's profits. Moreover, firms can employ research factors to learn about a commodity's production process without producing it commercially.[5] Consequently, firms can separate learning decisions from production decisions.

This separation implies that firms only consider period t's profits when they decide to begin or cease production. These decisions, therefore, can and will be formulated in terms of cost curves. A firm begins or ceases production according to whether price exceeds or is less than its minimum average cost (excluding *R&D* expenditures).

In addition to the firm behavior outlined above, assume the innovator (a domestic firm) maintains a monopoly position for b periods, i.e., barriers to entry disappear after period b. When the innovator loses his monopoly position, let the number of entering firms be sufficiently large that price is taken as given. Then, prior to period b, the innovator produces in the domestic country to exploit his learning advantage (Assumptions c, d, and e) and maximizes profits by equating marginal revenue to marginal cost.

In period b+1, due to their learning advantage (Assumptions c, d, and e), only domestic firms enter the market. All firms equate marginal cost to price and entrance ceases when price equals the marginal firm's minimum average cost. Since all firms entering in this period are marginal firms,[6] price is determined by their minimum average cost. Measuring profits exclusive of *R&D* expenditures, the firm that first entered the market earns a positive profit, while all other domestic firms' profits are zero. Subsequent periods can be described analogously; intramarginal firms earn positive profits, while marginal firms (those just entering the market) earn a zero profit.

Eventually, since the knowledge function is bounded and the foreign country has a static advantage, domestic firms will begin to shift their production to the foreign country. In other words, eventually the advantage of producing in the foreign country will outweigh cost reductions obtained by remaining in the domestic country to exploit the learning advantage. If the United States is the domestic country, therefore, this formulation would explain why comparative advantage often eventually shifts away from the United States. This in turn would explain why *U.S.* firms establish foreign subsidiaries.

II. The *U.S.* Learning Advantage

To identify the United States with the domestic country and apply the theory formulated above, one must show that *U.S.* firms have a learning advantage, and

[5] Commercial production begins when a firm has some knowledge of its product's production process and knows that there is a demand for its product. Prior to this, firms may and probably will engage in pilot-plant production (exploratory, small-scale production).

[6] Because of cumulative variables in knowledge functions, marginal domestic firms are those that have most recently entered the market.

that this advantage can be realized only by producing in the United States. The *U.S.* learning advantage results from the high marginal products of learning resources (scientists and engineers) in *U.S.* firms relative to other countries.[7] First, scientists and engineers in *U.S.* industry are evenly distributed between production (37 percent) and *R&D* (35 percent), with a significant percentage in management and administration (8 percent).[8] Comparable data for other countries have been found for Greece, the United Kingdom, and Canada. With the exception of Canada,[9] the distribution of scientists and engineers is skewed in these foreign countries relative to the United States. In Greece, 91 percent of the scientists and engineers are engaged in production,[10] while in the United Kingdom, the ratio of research to production scientists is five times that in the United States.[11]

Additional country data are necessary to determine to what extent Canada is an exception in having an even distribution of scientists and engineers throughout all employment levels of Canadian firms. One can conclude that the distribution of scientists and engineers throughout all employment levels in *U.S.* firms has contributed to their high marginal product in three ways:[12] by facilitating the ability of plant managers to transform *R&D* output into production; by increasing the ability of executives to evaluate research activity; and by increasing communication within the firm.

For instance, according to Robert Kornhauser:

> . . . the chairman of the board at American Brake Shoe Company asserts that the research department might have failed many times to get a new product into manufacture if there had not been men with an engineering background in management. [p. 176]

The above discussion dealt with integrating research personnel into the firm to increase the effective utilization of research Second, and equally important, is integrating management into research departments. A study of 405 research laboratories in the United States found that the research director usually had scientific training (Kornhauser, p. 176). In an intercountry comparison of *R&D*, the Organization for European Economic Cooperation (*OEEC*), stresses the importance of such training:

> If applied research programmes are to be effective and research organizations make their full contribution, they must be soundly and well administered. More attention should be given, as in the United States, to the training of research directors combining the qualities of the scientist and the manager. [p. 70]

Third, the marginal product of scientists and engineers is also higher in the United States because of the close cooperation between the university, government, and industry. The *OEEC* study (p. 70), found that this cooperation was greater in the United States than in Europe.

Finally, *U.S.* firms provide a very favorable environment for research and development. Industrial laboratories have replaced amateur inventors and unplanned individual innovation everywhere, creating potential conflicts within the firm between scientists and management. Kornhauser has analyzed these conflicts and has shown that *U.S.* firms have been very successful in resolving them, thereby creating a favorable environment for *R&D* in *U.S.* industry. A lack of data prohibits intercountry comparisons, but of eleven countries studied by the Organization for European Cooperation and Development

[7] Scientists and engineers are engaged not only in developing and improving production processes but also in developing new products. Consequently, if their marginal product is higher in *U.S.* firms, these firms must have an advantage in innovating in addition to learning.

[8] See *OECD*, p. 174. The importance of this even distribution is shown below.

[9] *OECD*, p. 152.

[10] See *OECD*, p. 152.

[11] See Merton J. Peck, pp. 454–86.

[12] See Peck, pp. 454–86.

(*OECD*) in 1963, the United States had the highest percentage (74 percent) of scientists and engineers in industry (see *OECD*, p. 117). This suggests that *U.S.* industry may provide a more favorable environment for these learning factors.

To apply the theory in Section I to the United States, one must show not only that *U.S.* firms have a learning advantage, but also that this advantage is limited to the geographical confines of the United States. Preceding arguments have only shown that *U.S.* firms have a learning advantage. One can show, however, that this advantage is best realized by producing in this country.

First, excellent communication between the university, government, and industry was listed as an important factor above, and this is clearly confined to the United States. Second, most of the *R&D* facilities of *U.S.* firms are located in this country, which is due partly to the excellent communication just mentioned. Third, the bulk of existing production facilities of *U.S.* firms is located in this country. Many innovations consist of modifications and improvements of existing products and development of related products. Existing production facilities, therefore, are both suitable and useful in the early production stages. This latter factor diminishes in importance as *U.S.* firms establish production facilities abroad. Nevertheless, it is evident from this discussion that the *U.S.* learning advantage is currently confined to this country.

III. Testing the Theory

The Model

To test this theory, a model has been formulated to explain the behavior of a cross-section of *U.S.* firms over time. The model contains four equations that explain applied research expenditures, new product innovation, research expenditures for process improvement, and the difference between foreign subsidiary sales and domestic sales. Since data have been found only to estimate the last equation, which tests the theory, only this equation will be discussed and estimated.[13]

The microtheory formulated in Section I requires one to specify this equation for an individual commodity:

$$
\begin{aligned}
& FS_{ijt} - DS_{ijt} \\
(1) \quad & = \overset{1}{c}_{ij}\cdot(FM_j) + h\cdot(F_{ij} - \hat{S}_{ij}(RD_{ijt-1})) \\
& \quad \cdot TS_{ijt-1} + \overset{2}{c}_{ij}\cdot(M_j) + u_{ijt}
\end{aligned}
$$

In equation (1), subscripts i, j, and t refer to firm i, product j, and time t. The variable FS_{ijt} denotes foreign subsidiary sales, DS_{ijt} denotes domestic sales. The variable FM_j is a dummy for product j that equals one in the first year that product j is marketed and zero otherwise. The variable F_{ij}, which is less than one, is the value of the knowledge function, S, when domestic firms switch production to the foreign country. In other words, it is the value of S when the foreign static advantage outweighs the domestic learning advantage. The function $\hat{S}_{ij}(RD_{ijt-1})$ is firm i's estimate of its knowledge function for product j.[14] The variable TS_{ijt-1} denotes total sales (domestic plus foreign). The variable M_j is a dummy for product j that equals zero in the first that year product j is marketed and one otherwise. The variable u_{ijt} is the disturbance term, and c_{ij}^1, c_{ij}^2, and h are negative constants.

Because of the nature of the dummy variables, FM_j and M_j, and except for the disturbance, c_i^1 is the only nonzero term in

[13] Even though only one equation will be estimated, the estimation procedure will take into account that it is part of a simultaneous equation model. For a complete discussion of equations in the model *not* discussed here, see my Ph.D. dissertation.

[14] Assuming firms attribute all learning to *R&D* (learning by doing is exogenous and relatively unimportant), $\hat{S}_{ij}$ depends only on cumulative *R&D*. Further assuming that firms decide in period t−1 whether or not to switch production locations, $\hat{S}_{ij}$ depends on RD_{ijt-1}.

the equation when product j is first produced. Similarly, c^2_{ij} is the only nonzero term when $\hat{S}_{ij}=F_{ij}$. Consequently, $-c^1_{ij}$ equals domestic sales when product j is first produced, while $-c^2_{ij}$ is the level of domestic sales immediately prior to the establishment of a foreign subsidiary.

In other years, the explanatory term is $h\cdot(F_{ij}-\hat{S}_{ij})$, which is directly related to the theory in Section I. Since h is negative, domestic production expands when domestic firms conclude that their opportunities for learning are sufficiently large ($\hat{S}_{ij}<F_{ij}$) to justify producing all of product j in the domestic country. When these firms believe that additional opportunities for learning are small ($\hat{S}_{ij}>F_{ij}$), they begin to shift production of product j to the foreign country. When this occurs, foreign subsidiary sales expand.[15]

The size of period t's domestic or foreign subsidiary sales depends not only on the knowledge function, but also on consumers' acceptance of the product. The term $F_{ij}-\hat{S}_{ij}(RD_{ijt-1})$ is multiplied, therefore, by the firm's sales of product j in period $t-1$.[16]

To test this equation, since data were unavailable for individual products, it was aggregated over all products:[17]

$$(FS_{it}-DS_{it}) = h\cdot\sum_{j=1}^{P_{it}}[(F_{ij}-\hat{S}_{ij}(RD_{ijt-1}))\cdot TS_{ijt-1}] + \sum_{j=1}^{N_{it}}(c^1_{ij}) + \sum_{j=1}^{(P_{it}-N_{it})}(c^2_{ij}) + \sum_{j=1}^{P_{it}} u_{ijt} \tag{2}$$

In equation (2) P_{it} is the number of products produced and N_{it} the number introduced by firm i at time t. Because data have not been obtained for most of the variables in equation (2), the following assumptions are made to transform (2) into a form in which it can be estimated:

(i) In choosing research factors, re_{ijt}, assuming the shortest horizon possible yields:[18]

$$re_{ijt} = \hat{E}_{st}\cdot\hat{T}S_{ijt+1}\cdot d(1) + e_{ijt}$$

where $\hat{E}_s$, denotes the firm's estimate of E_s, the elasticity of learning with respect to learning factors (re_{ijt}); $\hat{T}S_{ijt+1}$ denotes the firm's estimate (made at time t) of total sales at time t+1; $d(1)=1/(1+r)$ is a constant discount factor; e_{ijt} is a disturbance term. Further assume:

$$\hat{E}_{st} = a(1-\hat{S}_{ij}(RD_{ijt})), \quad a>0\text{[19]}$$

$$\hat{T}S_{ijt+1} = (1+g)TS_{ijt}, \quad g>0\text{[20]}$$

[15] One might prefer to employ real rather than nominal sales data. In the pharmaceutical industry, which will be used to test the theory, prices have remained relatively stable. Consequently, nominal sales are approximately equal to real sales.

[16] This is particularly important in the pharmaceutical industry, because doctors and other purchasing agents do not have time to evaluate each of the many drugs introduced each year; to a large extent they rely on information distributed by drug firms. That firm, *ceteris paribus*, which has been most successful in distributing such information will have the greater sales in period t. Let TS_{ijt-1} measure the success with which this information has been disseminated.

[17] This equation was derived to explain shifts in comparative advantage. Since production costs are low in the pharmaceutical industrial, tariffs and/or tax differentials may be important in determining the location of foreign subsidiaries of *U.S.* firms. In this event comparative advantage, as traditionally defined, would not necessarily shift to the foreign country. Nevertheless, the theory formulated here still applies; *U.S.* firms establish foreign subsidiaries when tariff and/or tax considerations dominate future opportunities for learning.

[18] Since new-product demand is uncertain, a firm may adopt the shortest planning horizon possible.

[19] This specification states that the elasticity, E, depends linearly on the distance of S from its upper bound; it is largest when learning first begins and subsequently declines. This corresponds to the assumption that learning consists initially of major breakthroughs, which are followed by refinements. When learning ceases ($S\approx 1$), E is zero as one would intuitively expect. It is interesting to note that those who argue that *R&D* expenditures bear a constant relation to sales may be implicitly assuming that this elasticity is constant.

[20] Since period t+1's sales of product j are unknown at time t, the firm must estimate them. One could argue that estimated sales in period t+1 are a distributed lag function of past sales. This possibility was not explored because of the small data sample. Instead, it was simply assumed that sales are expected to grow by some constant proportion, g.

(ii) A firm's total sales depend on prices, the "newness of its product," etc., but not on the switching point F_{ij}. Consequently, F_{ij} and TS_{ijt-1} are uncorrelated, implying that

$$\left[\sum_j F_{ij}/P_{it}\right]\cdot \sum_j TS_{ijt} \equiv \overline{F}_{it}\cdot TS_{it-1}$$

is an unbiased estimate of

$$\sum_j F_{ij}\cdot TS_{ijt-1}$$

Assuming $\overline{F}_{it}\approx F$, i.e., F_{it} is approximately constant:

$$\sum_j F_{ij}\cdot TS_{ijt-1} \approx F\cdot TS_{it-1}$$

(iii) Define $\bar{c}^1_{it}$ and $\bar{c}^2_{it}$ as:

$$\bar{c}^1_{it} \equiv \sum_{j=1}^{N_{it}} (c^1_{ij}/N_{it});$$

$$\bar{c}^2_{it} \equiv \sum_{j=1}^{(P_{it}-N_{it})} (c^2_{ij}/(P_{it}-N_{it}))$$

Assuming $\bar{c}^1_{it}\approx c^1$ and $\bar{c}^2_{it}\approx c^2$, i.e., these means are approximately constant, yields:

$$\sum_j c^1_{ij} \approx c^1\cdot N_{it};\quad \sum_j c^2_{ij} \approx c^2\cdot(P_{it}-N_{it})$$

Then, substituting (i), (ii), and (iii) into equation (2) yields an estimatable equation:

$$(3)\quad \begin{aligned}(FS_t - DS_{it}) &= b_1 re_{it-1} + b_2 TS_{ijt-1}\\ &+ b_3 N_{it} + b_4(P_{it}-N_{it}) + w_{it}\end{aligned}$$

In equation (3):

$b_1 \equiv h/(a\cdot d(1)) < 0$ since $h<0$ and $a\cdot d(1)>0$

$b_2 \equiv -h(1-F) > 0$ since $h<0$ and $F<1$

$b_3 \equiv c^1,\ c^1<0;\quad b_4\equiv c^2,\ c^2<0$

$$w_{it} = \sum_j u_{itj} + \sum_j v_{ijt-1},$$

$$v_{ijt} \equiv [-1/(a\cdot d(1))]e_{ijt}$$

and

$$j = 1, \ldots, P_{it}$$

The Data

Complete data were obtained for three nonproprietary drug firms for the period 1962–68 and for an additional firm for the period 1964–68. The sample of twenty-six observations was obtained by pooling a cross-section of these firms.

A National Science Foundation questionnaire (*NSF* #1) was used to determine domestic sales. These were defined to be sales of *U.S.* firms located in the United States. Total sales were obtained from the Compustat Tape (Standard Statistics Co., Inc.). Foreign subsidiary sales of each drug firm were then derived by subtracting domestic sales from total sales.

R&D data for these firms were also obtained from the *NSF* questionnaire. Ideally one would want *R&D* expenditures allocated to process improvement. The *NSF* questionnaire classified *R&D* expenditures as basic, applied, or development. The latter appears to correspond closely to expenditures for process improvement. The National Science Foundation has defined these expenditures to include:

> The design and operation of pilot plants or semiworks plants so long as the principal purposes are to obtain experience and to compile engineering and other data to be used in evaluating hypotheses, in writing product formulas, or in establishing finished product specifications, in designing special equipment and structures required by a process and in preparing operating instructions or manuals. [#2(1968), p. 137]

Unfortunately, in addition to process improvement, these data include expenditures for "writing product formulas" and "establishing finished product specifications." In other words, expenditures for testing new drugs are included in this category. Furthermore, due to more re-

strictive drug legislation, which became effective in 1963, these testing costs have increased in the sample period.[21]

Although data on testing costs were unavailable, they can be estimated by making several assumptions. Paul de Haen classifies New Drugs as new single ingredients, duplicate single ingredients, combination products, and new dosage forms. Most testing costs are undoubtedly incurred for the first of these, new single ingredients. Assume further that the cost per product changes only in response to changes in drug legislation. Denote α_{ij}^1 and α_{ij}^2 as the costs of testing new drugs in the periods 1960–62 and 1963–68, respectively. Then, since α_{ij}^1 and α_{ij}^2 are fixed costs, re_{it}, firm i's expenditure for process improvement at time t, is given by equation (4):

$$re_{it} = re'_{it} - [\bar{\alpha}^1_{it}\cdot(d_1\cdot NS_{it}) + \bar{\alpha}^2_{it}\cdot(d_2\cdot NS_{it})] \tag{4}$$

In equation (4) re'_{it} is firm i's development expenditures, defined according to the *NSF* definition.

$$\bar{\alpha}^1_{it} \equiv \sum_j \{\alpha^1_{ij}/N_{it}\}; \quad \bar{\alpha}^2_{it} \equiv \sum_j \{\alpha^2_{ij}/N_{it}\}$$

d_1 and d_2 are dummy variables corresponding to the periods 1960–62 and 1963–68, respectively. The term NS_{it} denotes the number of new single ingredients firm i introduced at time t.[22] Assume that $\bar{\alpha}^1_{it}$ and $\bar{\alpha}^2_{it}$ are constant, i.e., $\bar{\alpha}^1_{it}=\bar{\alpha}^1$ and $\bar{\alpha}^2_{it}=\bar{\alpha}^2$. Then, substituting equation (4) into (3), and letting $b_5\equiv -\bar{\alpha}^1\cdot b_1$ and $b_6\equiv -\bar{\alpha}^2\cdot b_1$, yields equation (5) which will be estimated:

$$\begin{aligned}(FS_{it} - DS_{it}) &= b_1 re'_{it-1} + b_2 TS_{it-1} + b_3 N_{it} \\ &\quad + b_4(P_{it} - N_{it}) + b_5(d_1 NS_{it-1}) \\ &\quad + b_6(d_2 NS_{it-1}) + w_{it}\end{aligned} \tag{5}$$

Additional data on applied research expenditures on the industry level were also obtained from *NSF* (#2). This variable does not appear in equation (5), but is one of the exogenous variables in the model that will be used to estimate it. Finally, the number of drugs produced and introduced by each firm, each year, was obtained from Medical Economics, Inc., *Physicians' Desk Reference to Pharmaceutical Specialties and Biologicals.*

The Estimation Method

To estimate (5), note that the disturbance term, w_{it}, equals

$$\sum_{j=1}^{P_{it}} u_{ijt} + \sum_{j=1}^{P_{it}} v_{ijt-1}$$

Assume that each term can be decomposed as follows:

$$u_{ij} \equiv \overset{I}{u}_{ij} + \overset{II}{u}_{ijt};\ v_{ijt} \equiv \overset{I}{v}_{ij} - \overset{II}{v}_{ijt}$$

where u^I_{ij} and v^I_{ij} are time-invariant firm and/or product effects.[23] Further assume that u^I_{ij}, u^{II}_{ijt}, v^I_{ij}, and v^{II}_{ijt} are mutually independent and that u^{II}_{ijt} and v^{III}_{ijt} are serially uncorrelated. Then the variance-covariance matrix of the disturbance, Ω, is block-diagonal. A typical element of block 1, σ_{ij}, would be:[24]

i) $P_{1j}\cdot\sigma^2$ if i=j, where P_{1j} is the number of products firm 1 produces at time j; $\sigma^2\equiv$variance of w_{it}.

[21] This legislation consists of the Kefauver-Harris Drug Amendments and Investigational Drug Regulations.

[22] See de Haen (parts II and IV). I would like to thank him for supplying these data.

[23] This is similar to Pietro Balestra and Marc Nerlove's formulation, pp. 585–612. The disturbance is specified in this manner first because preliminary results indicate that autocorrelation probably exists. Second, it seems likely that there are differences among firms or products that are captured in the disturbance term.

[24] This specification implicitly assumes that for a new product there is at least a one-period lag between research expenditures for process improvement and the new product's introduction.

ii) $C_I^{ij} \cdot \sigma_I^2$ if $i \neq j$, where C_I^{ij} is the number of products produced in period i that were also produced in period j by firm 1; $\sigma_I^2 \equiv$ variance of $u_{ij}^I + v_{ij}^I$.

The assumptions above imply that Ω is not of the form $\sigma^2 \cdot I$. Moreover, TS_{it-1} and re_{it-1}^i are correlated with the disturbance. To estimate (5), therefore, a combination of Generalized-Least Squares (*GLS*) and Instrumental Variable (*IV*) estimation procedures were employed.[25]

IV: Conclusions

Estimation Results

The estimation results are shown in Table 1 for equation (5). Since $\sigma^2 \equiv \sigma_I^2 + \sigma_{II}^2$ and σ^2, σ_I^2, and σ_{II}^2 are all positive, $0 < \sigma_I^2/\sigma^2 \leq 1$. Since σ_I^2 refers to time invariant effects, they are present if σ_I^2/σ^2 equals one and absent if it equals zero. Since the estimate of this ratio was .83, these effects are probably present.

TABLE 1—ESTIMATION RESULTS[a]

b	Expected Sign	Estimate of b: $\hat{b}$	Standard Deviation	t-Statistic	F	σ_I^2/σ^2
b_1	−	−.0168*	.0017	−9.878		
b_2	+	.2005*	.0674	2.972		
b_3	−	1.3352	1.3221	1.010	9.18	.83
b_4	−	−.3578***	.2210	−1.619		
b_5	+	−13.2723	29.3280	.452		
b_6	+	8.9046**	4.9689	1.792		

[a] *, **, and *** indicate a 1, 5, or a 10 percent significance level in a one-tail test.

An F-ratio was computed to test the null hypothesis that all of the coefficients equal zero.[26] Since this ratio equals 9.18, the null hypothesis was rejected at the 1 percent level, implying that a significant relation exists between dependent and independent variables.

Except for $\hat{b}_3$ and $\hat{b}_5$, which were insignificantly different from zero, all coefficients had the expected sign. Since $-b_3$ equals the initial level of domestic sales, $\hat{b}_3$ should have been negative, but was positive. This may be due to misspecifying b_3, which was assumed to be constant and equal for all firms. These assumptions are questionable, but data did not permit an alternative specification.

Fourth, since $-b_5/b_1$ equals the average level of expenditures for testing a new drug from 1960 to 1962 and b_1 is negative, $\hat{b}_5$ should have been positive, but was negative. This unsatisfactory estimate of b_5 might indicate that development expenditures were inadequately corrected for new product testing costs.[27]

Finally, all other coefficients were significantly different from zero and had the expected sign. The estimates of $\hat{b}_1$ and $\hat{b}_2$ are of particular importance; they are the most significant estimates (they have the highest t-statistics) and are directly related to the theory in Section I. Since b_1 equals $h/(a \cdot d(1))$ and b_2 equals $-h \cdot (1-F)$, a negative value for $\hat{b}_1(a \cdot d(1)) > 0$ and a positive value for b_2 $(F<1)$ both imply that h is negative. But, h is the coefficient of the term $((F_{ij} - S_{ij}(RD_{ijt-1}))$.

[25] Write equation (5) as:

$$Y = XB + W \tag{5a}$$

Let T be a matrix such that $T'T = \Omega^{-1}$. Then, multiplying (5a) by T yields:

$$(TY) = (TX)B + (TW) \tag{5b}$$

Since the variance-covariance matrix of TW is of the form $\sigma^2 \cdot I$, it seems reasonable to apply instrumental variables to (5b). Instruments for TS_{it-1} and re_{it-1}^i were chosen by regressing these variables on all of the model's exogenous variables in the model (see my doctoral dissertation for a discussion of model's equations and exogenous variables). Letting Z be a matrix of instruments, TZ is the matrix of transformed instruments. Then, multiplying (5b) by $(TZ)'$ and setting $(TZ)'(TW)$ equal to zero, one can solve for $\hat{B}$, the *GLS-IV* estimate of B:

$$\hat{B} = (Z'\hat{\Omega}^{-1}X)^{-1}Z'\hat{\Omega}^{-1}Y$$

[26] $$F = \frac{[Y'\Omega^{-1}Y - (Y - X\hat{B})'\Omega^{-1}(Y - X\hat{B})]/6}{(Y - X\hat{B})'\hat{\Omega}^{-1}(Y - X\hat{B})/20},$$

distributed $F(6, 20)$.

[27] Since d_1 was nonzero in only two years of the sample, there were only a few observations on $(d_1 \cdot NS_{it-1})$. This could explain the low t-statistic for b_6.

Consequently, a negative value for h means that foreign subsidiary sales increase as opportunities for additional learning by *U.S.* pharmaceutical firms decline. This accords fully with what the theory of Section I would predict.

Limitations of Study

The sample size (twenty-six observations) limited the manner in which the theory was tested. Aside from questionable large-sample properties of the estimates, several interesting hypotheses could not be tested. For example, the speed with which *U.S.* firms locate abroad, the size of testing costs, and the initial and maximum level of domestic sales may vary among firms. In other words, the equation tested may be different for each firm. To test this hypothesis, however, one would have to estimate this equation for each firm. Since there were at most seven observations per firm, this was not feasible.

Another limitation was imposed by the lack of data on *U.S.* foreign subsidiary sales in each country. Had country data been available, one could have examined the special characteristics of each country (or group of countries in the case of the Common Market).

A third limitation was imposed by the lack of data on research expenditures for process improvement by pharmaceutical firms. As stated in Section III, these were estimated by subtracting an estimate of new product testing costs from development expenditures.

Finally, because individual product data were unavailable for all variables, it was difficult to interpret the estimation results. The theory formulated here pertains to an individual product, implying that the equation developed to test the theory should also refer to an individual product. Such an equation was formulated, but had to be aggregated because individual product data were unavailable. Unfortunately, because of the aggregation, each of the b_i in this equation is really the composite of several coefficients. For example, $b_1 \equiv [h/(a \cdot d(1))]$. Consequently, it was only possible to infer signs rather than actual values of certain key coefficients (i.e., h).

Conclusions

The theory formulated in Section I differs from previous formulations by recognizing the role of research and development. To a large extent firms allocate *R&D* to learn about their production processes, the cumulative effects of which are formalized in a knowledge function.

A simple formulation of this function enables one to show both why a new product is often first developed and produced in the United States, and why this country often loses its initial advantage over time. Product cycle theories also attempt to explain this behavior.[28] Such theories have made important contributions to trade theory, but exhibit several deficiencies. First, they often claim that firms produce initially in this country to minimize demand uncertainties; but this only explains why a new product is marketed, not why it is produced in this country. The theory developed here is based on supply uncertainties. As shown in Section II, this enables one to explain why new products are often produced in the United States.

Second, many product cycle theories attempt to explain shifts in comparative advantage away from the United States in terms of "technology stabilization." The knowledge function both clarifies and explains this process. A "stable technology" is one in which the stock of knowledge has become constant, because the marginal

[28] Product cycle theories were developed by Gary Hufbauer, pp. 1–165; Seev Hirsch (1965), pp. 92–97, (1967), pp. 1–133; William Gruber, Dileep Mehta, and Raymond Vernon, pp. 10–37; Michael Posner, pp. 232–41; Louis Wells, Jr., pp. 152–62; and Vernon, pp. 190–207.

product of cumulative *R&D* approaches zero. Note that comparative advantage need not shift when the stock of knowledge becomes constant, but rather when the foreign static advantage dominates the *U.S.* learning advantage.

Third, with little explanation, product cycle theories often claim that a low wage country has the static advantage described above. In other words, comparative advantage eventually shifts to such a country. This can now be explained by making three assumptions. First, investment by *U.S.* firms in foreign subsidiaries suggests that capital is mobile. Second, as argued in Section II, *R&D* is relatively immobile. Third, assume that labor is also relatively immobile. Then, since capital mobility implies that inter-country factor payments to capital converge, *R&D* and labor are the main factors in determining comparative advantage. Early in the production process, when learning is important, the *U.S.* has a comparative advantage in producing a new product. When learning becomes less important, it is then not surprising that comparative advantage should shift to a low wage country.

In conclusion, this theory argues that the dynamics of comparative advantage are due to learning. It differs from previous formulations, because it is based on those supply conditions essential to developing the production process for a new product.[29]

[29] This formulation pertains to a single, new commodity. In a multi-commodity world, one would expect the country with a learning advantage to continually shift its production to new commodities, because its greatest advantage will always be early in the learning process. If the United States has a learning advantage, then it probably has a comparative advantage in producing new products, which would imply that *U.S.* exports have a high *R&D* content. This is precisely the case, for the five industries representing only 39 percent of total *U.S.* sales account for: 1) 72 percent of *U.S.* exports; 2) 89 percent of *U.S.* research and development expenditures. (See Gruber, Mehta, and Vernon, pp. 22–23.)

REFERENCES

K. J. Arrow, "The Economic Implications of Learning by Doing," *Rev. Econ. Stud.*, June 1962, *39*, 155–73.

P. Balestra and M. Nerlove, "Pooling Cross Section and Time Series Data in the Estimation of a Dynamic Model: The Demand for Natural Gas," *Econometrica*, June 1966, *34*, 385–612.

P. de Haen, *De Haen Nonproprietary Name Index with Therapeutic Guide*, New York 1969.

W. Gruber, D. Mehta, and R. Vernon, "The R and D Factor in International Trade and International Investment of U.S. Industries," *J. Polit. Econ.*, Feb. 1967, *75*, 20–37.

S. Hirsch, *Location of Industry and International Competitiveness*, Oxford 1967.

———, "The U.S. Electronics Industry in International Trade," *Nat. Inst. Econ. Rev.*, Nov. 1965, 92–97.

G. C. Hufbauer, *Synthetic Materials cnd the Theory of International Trade*, London 1966.

R. W. Klein, "A Dynamic Theory of Comparative Advantage," unpublished doctoral dissertation, Yale Univ. 1970.

R. Kornhauser, *Scientists in Industry*, Berkeley 1967.

A. Madansky, "On the Efficiency of Three-Stage Least Squares Estimation," *Econometrica*, Jan.-Apr. 1964, *32*, 51–56.

M. J. Peck, "Science and Technology," in F. Caves and Associates, eds., *Britain's Economic Prospects*, Washington 1958.

M. V. Posner, "International Trade and Technical Change," *Oxford Econ. Pap.*, Oct. 1961, *13*, 323–41.

R. Vernon, "International Investment and International Trade in the Product Cycle," *Quart. J. Econ.*, May 1966, *80*, 190–207.

L. T. Wells, "Test of a Product Cycle Model of International Trade: U.S. Exports of Consumer Durables," *Quart. J. Econ.*, Feb. 1969, *83*, 152–62.

Medical Economics, Inc., *Physicians' Desk Reference to Pharmaceutical Specialities and Biologicals*, New Jersey 1959–60.

National Science Foundation (#1), "Form RD-1, Survey of Industrial Research and

Development," unpublished questionnaire, Washington 1960–68.

———, (#2), *Research and Development in Industry,* Washington 1962–69.

Organization for European Cooperation and Development (*OECD*), *Resources of Scientific and Technical Personnel in the OECD Area,* Paris 1963.

Organization for European Economic Cooperation (*OEEC*), *The Organization of Applied Research in Europe,* Paris 1954.

Standard Statistics Co., Inc., *Compustat Tape* (computer tape), New York 1968.

U.S. Department of Commerce, *U.S. Foreign Trade, a Five-Year Outlook with Recommendations for Action,* Washington 1969.

A True Price Index When The Consumer Saves

By MALCOLM GALATIN*

The conventional economic theory of price or cost of living index numbers defines a true price index (*TPI*) as the ratio of two volumes of expenditure which are the minimum needed by a consumer to attain a reference level of utility in two price situations. When the two price situations are designated 0 and 1, respectively,[1] the reference level of utility is generally chosen to be that attained by the consumer in situation 0 or in situation 1. Once these *TPI*s have been defined, they may be related to measurable price indexes, and we have the well-known results that when the reference level of utility is that attained in period 1, the Paasche price index is a lower bound of a *TPI*, and when the reference level of utility is that attained in period 0, the Laspeyres' price index is an upper bound to a *TPI*.[2]

If the consumer's income, expenditure, and utility function are assumed to be the same before and after a price change, the concept of a *TPI* is closely related to that of a conceptual income compensation that must be added to (or subtracted from) his income to leave him as well off after the change in prices as he was before.[3] Thus, if we imagine a consumer beginning in a situation with period 0 prices and an income of M_0, he would be as well off after a price change to prices of period 1 when C_0, given by (1), is added to (or subtracted from) M_0.

$$C_0 = M_0(TPI_0 - 1) \tag{1}$$

The value TPI_0 is the *TPI* measured with the attained period 0 utility as a reference level. Similarly, if the consumer is imagined to begin in a situation 1 with an income of M_1, and then prices change to those of situation 0, the consumer would be as well off before as after the price change, if his income, M_1, is compensated by C_1, given by (2).

$$C_1 = M_1\left(\frac{1}{TPI_1} - 1\right) \tag{2}$$

The value TPI_1 is the *TPI* measured with the attained utility level of period 1 as the reference level of utility.

The fact that a consumer saves part of his income has not been explicitly considered in the literature dealing with the economic theory of index numbers, and it is the purpose of this paper to extend the theory to take account of this fact. Superficially the existence of saving might seem to pose no problem. If saving is considered as a good, then it has a price, and it might be thought that the existing theory is adequate to deal with it. For example, assume that a consumer with a money income of

* Assistant professor of economics, City College of the City University of New York. An earlier version of this paper was read at the December 1969 meetings of the Econometric Society.

[1] Situation 1 may be thought of as the current period and situation 0 a past period.

[2] See for example Melville Ulmer, especially ch. 3.

[3] Here it is assumed that if the consumer is on the same indifference surface before and after the price change, he really would obtain the same satisfaction in both situations. If this assumption is not made, an alternative approach is to reinterpret the *TPI* in the manner suggested by Franklin Fisher and Karl Shell and, likewise, to reinterpret the meaning of the income compensation. (This reinterpretation does not alter the formal definitions of a *TPI* or the income compensation.)

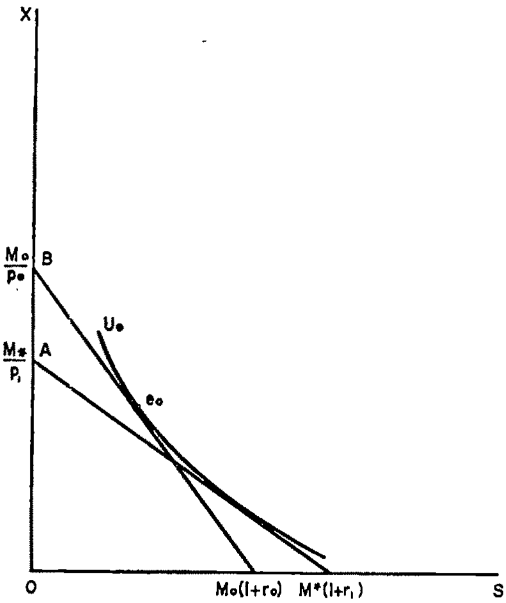


FIGURE 1

M_0 in period 0 has to decide how to allocate his income between one good X in amount x_0 and saving $S_0/(1+r_0)$, where r_0 is the market rate of interest. He has a utility function given by

$$U_0 = U(x_0, S_0) \tag{3}$$

and his budget constraint is

$$M_0 = p_0x_0 + S_0/(1 + r_0) \tag{4}$$

where p_0 is the unit price of X and $(1+r_0)^{-1}$ is the "price" per unit of saving owned at the end of period 0. The consumer would be in equilibrium at e_0 in Figure 1, attaining a utility level of U_0. Now, if in period 1 prices change to p_1 and $(1+r_1)^{-1}$, the minimum money income necessary to reach U_0 is M^*, and a *TPI* is defined as M^*/M_0 which is equivalent to $(OA\cdot p_1)/(OB\cdot p_0)$.

However, this analysis is inadequate because the consumer's saving decision is assumed to be made in a one period context and hence the reference level of utility is taken to be that of a single period for purposes of defining a *TPI*. Following the work of Irving Fisher, the modern view of the consumer's saving behavior is that he makes an intertemporal allocation of his expenditures to maximize an intertemporal utility function rather than a single period utility function.[4] It is this view of the consumer's saving decision which is accepted in this paper, and its implications for the measurement of the effect of price changes on a consumer who saves leads to the *TPI* which is discussed below.

In Section I, a basic model of the consumer's intertemporal decision making is constructed. In Section II, a *TPI* is defined, and then in Section III, the relationship of this *TPI* to the measurable Paasche and Laspeyres' price indexes is derived.

I. The Basic Model

We shall consider a consumer who has to allocate his expenditures between two periods, 0 and 1. He has an intertemporal utility function given by

$$W = W(U_0, U_1) \tag{5}$$

where

$$U_j = U(x_j) \quad \text{where} \quad j = 0, 1 \tag{6), (7}$$

is the utility level in period j. The utility function for each period is assumed to be the same and so are the n goods available for consumption. The term x_j is an n dimensional column vector of goods consumed in period j.

The consumer is assumed to know period 0 prices, p^0, with certainty at the beginning of period 0, and he expects the prices of these goods to be p_e^1 in period 1. He has these expectations at the beginning of period 0 and he acts as if he is certain that these prices will occur.[5] The terms p^0

[4] As another example of this approach, see the article by Menahem Yaari.

[5] Uncertainty may be introduced by assuming that at the beginning of period 0, the consumer has a subjective probability distribution for period 1 prices, and allocates his expenditures between the two periods to maximize the expected value of W. In this case, his

and p_e^1 are n dimensional row vectors. Further, it is assumed that he knows his money income M_0 and M_1 in the two periods and also r_0, the market rate of interest in period 0. We also assume that he is not concerned with periods after 1 and makes no allowance for a terminal stock of savings at the end of period 1.

The consumer is assumed to make his decisions at the beginning of period 0. Given p^0, p_e^1, M_0, M_1, r_0, and knowledge of (5), (6), and (7), he must then decide how to allocate his expenditure over periods 0 and 1. This results in a decision on how much he will save, S_0, in period 0.

Formally we may write out his process of decision making as follows: Find numbers S_0, U_0, U_1, and vectors x_j, $j=0, 1$ which maximize equation (5) subject to equations (6)–(9), and inequalities (10) and (11).

$$M_0 - S_0 = p^0 x_0 \tag{8}$$

$$M_1 + S_0(1 + r_0) = p_e^1 x_1 \tag{9}$$

$$-M_1/(1 + r_0) \leq S_0 \leq M_0 \tag{10}$$

$$x_j \geq 0, \qquad j = 0, 1 \tag{11}$$

The solution to this constrained maximization problem will yield the following: his planned saving $\overline{S}_0$;[6] the market baskets he plans to consume in periods 0 and 1, given by (12) and (13), respectively; his

$$\bar{x}_0 = (\bar{x}_{10}, \bar{x}_{20}, \ldots, \bar{x}_{n0})' \tag{12}$$

$$\bar{x}_1 = (\bar{x}_{11}, \bar{x}_{21}, \ldots, \bar{x}_{n1})' \tag{13}$$

planned single period utilities $\overline{U}_0$ and $\overline{U}_1$; and his planned intertemporal utility $\overline{W}$. The bar above a variable will be used to indicate a planned quantity of this variable decided on at the beginning of period 0. Also, it should be noted that constraint (10) allows for borrowing in period 0. Finally we define

$$\overline{E}_0 = M_0 - \overline{S}_0 \tag{14}$$

to be his planned expenditure in period 0, and

$$\overline{E}_1 = M_1 + \overline{S}_0(1 + r_0) \tag{15}$$

to be his planned expenditure in period 1.

II. The Definition of a True Price Index

There is a large degree of arbitrariness in the measurement of a conventional *TPI*, because many single period reference utility levels may be chosen. In the following definition of a *TPI* when the consumer saves a degree of arbitrariness remains. This stems not so much from the choice of a reference level of utility but rather from a choice of the period or periods in which the consumer's expenditure is assumed to be compensated for the price change. Since our basic model of saving is a closed two-period model, the effect of a price change is to move the consumer from the level of intertemporal utility that he would have attained had prices remained constant to the level of intertemporal utility he attains given actual period 1 prices. It is the intertemporal utility level that would be attained if prices remain constant which is taken as our reference utility level for measuring a *TPI*. This is an intertemporal level as opposed to the single period reference level of conventional index number theory. We shall not be concerned directly with the consumer's price forecasting ability, because his intertemporal expenditure allocation which partly depends on his price expectations, is made prior to his knowledge of period 1 prices, and is not affected by an actual change in prices. Now let us return to the situation of the consumer of the basic model at the beginning of period 1.

expenditures in the two periods will generally differ from his chosen expenditures in the basic model above. However, the argument in the remainder of the paper is not substantially changed when the basic model is replaced by this decision model with uncertain prices.

[6] $\overline{S}_0$ may be zero, but this does not affect the results of the paper.

In the beginning of period 1, the actual prices of period 1, $\hat{p}^1$, become known to the consumer. Let us assume that $\hat{p}^1$, p^0, and p_e^1 are three different price vectors.[7] This means that actual prices have changed and also that the consumer incorrectly forecast period 1 prices. His plans for period 0 have been realized, and his expenditure in period 1 is at the planned level of $\overline{E}_1$. However, given the actual prices of period 1, the consumer will purchase a market basket of $\hat{x}_1$ in period 1, and in general $\hat{x}_1 \neq \bar{x}_1$. Thus his actual utility level in period 1 will be $\hat{U}_1$ where

$$(16) \qquad \hat{U}_1 = U(\hat{x}_1)$$

and his actual level of intertemporal utility will be $\hat{W}$ where

$$(17) \qquad \hat{W} = W(\overline{U}_0, \hat{U}_1)$$

In general,[8] $\hat{U}_1 \neq \overline{U}_1$ and $\hat{W} \neq \overline{W}$. Hence the effect of an unanticipated price change is to change the consumer's single period utility in period 1 from that expected and has a corresponding effect on his intertemporal utility.

However, it is not the divergence of plans from realizations that we want to concentrate on, but rather the divergence of realizations from the situation that would have existed had prices remained constant. Assume now that $\hat{p}^1 = p^0$ and still that $p_e^1 \neq \hat{p}^1 = p^0$. In this case the consumer in period 1 with an expenditure of $\overline{E}_1$ would purchase a market basket $\tilde{x}_1$. In general $\tilde{x}_1 \neq \bar{x}_1$ and $\tilde{x}_1 \neq \hat{x}_1$. He would then attain a utility level in period 1 of $\tilde{U}_1$ where

$$(18) \qquad \tilde{U}_1 = U(\tilde{x}_1)$$

and an intertemporal utility level of $\tilde{W}$ where

$$(19) \qquad \tilde{W} = W(\overline{U}_0, \tilde{U}_1)$$

In general,[9] $\tilde{U}_1 \neq \overline{U}_1$, $\tilde{U}_1 \neq \hat{U}_1$, $\tilde{W} \neq \overline{W}$ and $\tilde{W} \neq \hat{W}$.

These relationships are shown in Figure 2 where the intertemporal utility function (5) is drawn in expenditure space with prices holding in periods 0 and 1 taken as parameters. Details of the construction of these intertemporal indifference curves are given in Appendix A. An intertemporal indifference curve is labeled by $W[p]$ where the actual level of intertemporal utility is shown by a symbol over the W and the prices holding in period 1 are shown by p. The price vector p may take three forms: p^0, implying that prices remain unaltered between periods 0 and 1; $\hat{p}^1$, the actual prices of period 1; and p_e^1, the prices expected in period 0 to hold in period 1. The intertemporal indifference curves in Figure 2 are drawn on the assumption that p^0, $\hat{p}^1$, and p_e^1 are three different price vectors.

The intertemporal budget constraint of the consumer is AB given by equation (20).

$$(20) \qquad E_0 = M_0 + (M_1 - E_1)/(1 + r_0)$$

In the beginning of period 0, he plans to be at C on AB spending $\overline{E}_0$ and $\overline{E}_1$ in periods 0 and 1, respectively, and attaining an intertemporal utility of $\overline{W}$ on $\overline{W}[p_e^1]$. Given his savings decision he is constrained to be at C, but because $p_e^1 \neq \hat{p}^1$, the level of intertemporal utility he actually attains is $\hat{W}$ on $\hat{W}[\hat{p}^1]$. If prices had remained constant between periods 0 and 1, his expenditure allocation at C resulting from his savings decision would have allowed him to achieve an intertemporal utility level of $\tilde{W}$ on $\tilde{W}[p^0]$.

This intertemporal utility level $\tilde{W}$ is taken as the reference level of utility for the purpose of defining a *TPI* when the consumer saves. If there had been no price change, he would have attained $\tilde{W}$. The effect of the price change is to give the con-

[7] See Appendix B where the implications of different relationships between $\hat{p}^1$, p^0, and p_e^1 are discussed.

[8] See Appendix B for the implications of exceptions to these "general" results.

[9] See Appendix B for exceptions to these relationships between the single and intertemporal utility levels.

sumer an intertemporal level of $\hat{W}$.

We now draw the intertemporal indifference curve $\tilde{W}[\hat{p}^1]$ in Figure 2. $\tilde{W}[\hat{p}^1]$ represents the level of intertemporal utility $\tilde{W}$, the same as the level of $\tilde{W}[p^0]$, but now the actual prices of period 1, $\hat{p}^1$, are the relevant parameters. Hence, if the consumer is to have his income adjusted to compensate for the price change in period 1, he must receive sufficient income to move from C to some point on $\tilde{W}[\hat{p}^1]$. This income compensation would enable the consumer to attain the same level of intertemporal utility he would have attained had prices remained unchanged.

Here is where the main degree of arbitrariness arises in defining a *TPI* when the consumer saves, for clearly this income adjustment could be made in numerous ways. For example, we could use as the compensated level of income the present value of that smallest income stream which would just be sufficient to move the consumer from C to $\tilde{W}[\hat{p}^1]$.[10] This income stream divided by his actual present value of expenditures would provide a *TPI*. Alternatively we might imagine a conceptual expenditure adjustment in period 0

[10] Where E_1 is discounted by $(1+r_0)$.

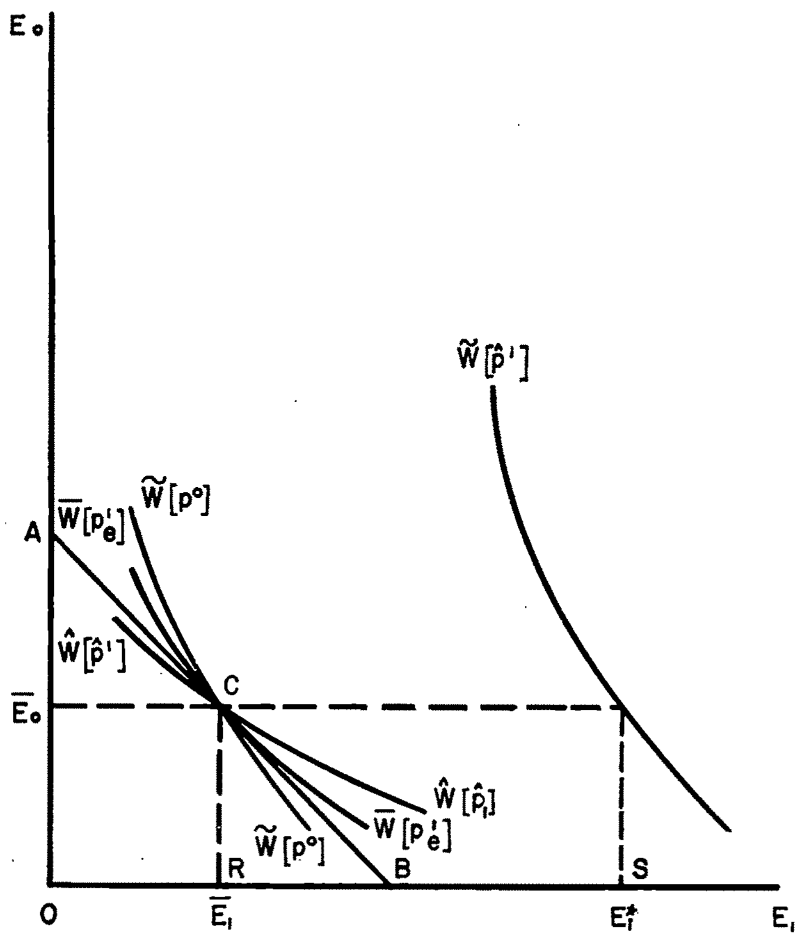


Figure 2

which is just sufficient to raise the consumer from C to some point on $\tilde{W}[\hat{p}^1]$.[11] This expenditure level divided by $\bar{E}_0$ would give a different TPI.

In this paper we shall assume that the compensation to his expenditure is made only in period 1. This is done for operational purposes, for once a price change has taken place the consumer cannot be compensated in period 0, but he can be compensated in period 1. Hence the minimum expenditure compensation in period 1 necessary to move the consumer from C to $\tilde{W}[\hat{p}^1]$ is $E_1^*-\bar{E}_1$, and we define a TPI as

$$TI = E_1^*/\bar{E}_1 \tag{21}$$

where E_1^* is the minimum expenditure the consumer needs in period 1 to achieve the level of intertemporal utility he would have achieved had prices remained constant, and $\bar{E}_1$ is his actual expenditure in period 1. Thus the situation of Figure 2 represents a true price increase for the consumer and the TI is given by OS/OR.

The effect of choosing period 1 for compensating the consumer's expenditure means that the measurement of a TPI when the consumer saves, based on an intertemporal reference utility level of $\tilde{W}$, is reduced to the measurement of a conventional TPI based on the single period reference utility level $\tilde{U}_1$. However, this TPI, TI defined in (21), is not the one that is usually measured. Rather $\bar{U}_0$ or $\hat{U}_1$ are taken as single period reference utility levels for measuring a TPI.

It has been assumed above that the price vectors p_e^1, p^0, and $\hat{p}^1$ are actually different from each other, and that the levels of intertemporal utility $\bar{W}$, $\tilde{W}$, and $\hat{W}$ are also different. In Appendix B the implications of alternative relationships between these prices and intertemporal utility levels are considered. It should be noted that a necessary condition for the TI to equal unity is that $\tilde{W}[\hat{p}^1]$ passes through C in Figure 2. This will happen when $p^0=\hat{p}^1$, a sufficient but not a necessary condition for the TI to equal unity. Further, the consumer's price forecasting ability is shown by how close $\bar{W}[p_e^1]$ is to $\hat{W}[\hat{p}^1]$ in Figure 2. If he correctly forecasts period 1 prices, these two intertemporal indifference curves are identical, represent the same level of intertemporal utility, and are tangent to AB at C. But no matter how proficient the consumer is in forecasting prices, the TI will only be unity when $\tilde{W}[\hat{p}^1]$ passes through C.

[11] This compensation would be added as a lump sum to $\bar{E}_0$.

III. The Relationship Between the True Price Index and Measurable Indexes

If we knew the form of the single period utility functions (6) and (7) and the intertemporal function (5), then the TI given by (21) could be measured exactly. Without knowledge of these functions, however, the TI may be related to measurable Paasche and Laspeyres' price indexes. As we would expect these results differ from the conventional ones because in effect our TI is a conventional TPI based on a single period utility level which is not generally used as the reference level of utility.

First we shall consider the relationship of the TI to the Paasche price index (PI). The compensated income in period 1 is E_1^* where

$$E_1^* = \hat{p}^1 x_1^* \tag{22}$$

The market basket, x_1^*, that the consumer would purchase in period 1 with his compensated expenditure is indifferent to $\bar{x}_1$ for him, since they would both yield him $\tilde{U}_1$ in period 1. Thus we may write

$$\bar{x}_1 I x_1^* \tag{23}$$

where I indicates indifference between market baskets. The PI is given by

$$PI = \bar{E}_1/E_1^+ \tag{24}$$

where

$$E_1^+ = \overset{0}{p}\hat{x}_1 \tag{25}$$

and

$$\overline{E}_1 = \overset{1}{\hat{p}}\bar{x}_1 = \overset{1}{p_e}\bar{x}_1 = \overset{0}{p}\bar{x}_1 \tag{26}$$

It should be remembered that with an expenditure of E_1^* and period 1 prices, the consumer would be observed to purchase x_1^*; with an expenditure of $\overline{E}_1$ and period 0 prices, the consumer would be observed to purchase $\bar{x}_1$; and with an expenditure of $\overline{E}_1$ and actual period 1 prices, the consumer is observed to purchase $\hat{x}_1$. However, E_1^+ is simply the cost of $\hat{x}_1$ at period 0 prices, and the consumer will not in general be observed to purchase $\hat{x}_1$ with an expenditure of E_1^+ and period 0 prices.

With this information the following results may be derived relating the TI to the measurable PI:[12]

THEOREM 1: When the $PI>1$, the $TI>1$

THEOREM 2: When the $TI<1$, the $PI<1$

THEOREM 3: When the $TI=1$, the $PI \leq 1$

THEOREM 4: When the $PI=1$, the $TI \geq 1$

The statements $PI=1$ and $TI=1$ in Theorems 3 and 4, respectively, hold only when $\hat{p}^1=p^0$. As an example we shall prove Theorem 1.

PROOF of Theorem 1:

When the $PI>1$, we have

$$\overline{E}_1 > E_1^+ \tag{27}$$

hence from (25) and (26)

$$p^0\bar{x}_1 > p^0\hat{x}_1 \tag{28}$$

which implies that

$$\bar{x}_1 RPT \hat{x}_1 \tag{29}$$

where RPT stands for "revealed preferred to" and we use the axioms of the theory of revealed preference.[13] Then from (23) and (29), we have

$$x_1^* RPT \hat{x}_1 \tag{30}$$

hence

$$\overset{1}{\hat{p}}\hat{x}_1 < \overset{1}{\hat{p}}\overset{*}{x_1} \tag{31}$$

hence

$$\overline{E}_1 < E_1^* \tag{32}$$

Combining (27) and (32), we have

$$E_1^* > \overline{E}_1 > E_1^+ \tag{33}$$

that is

$$E_1^*/\overline{E}_1 > 1 > (\overline{E}_1/E_1^+)^{-1} \tag{34}$$

hence

$$TI > 1 > 1/PI \tag{35}$$

therefore

$$TI > 1 \tag{36}$$

and Theorem 1 is proved. Notice that nothing is indicated about the relative size of the PI and the TI.

The proofs of Theorems 2 to 4 and the remaining one of this section will be provided by the author on request. They are similar to that of Theorem 1. In relating the TI to the PI, we are able to rank the PI and the TI only in the case when the $PI=1$. In this situation, the PI will be a lower bound on the TI if $\hat{p}^1 \neq p^0$, and equal to the TI if $\hat{p}^1=p^0$.

We can get additional information in

[12] The consumer is assumed to strongly order his preferences, i.e., he will always purchase a unique market basket when he has a given money expenditure and faces the same market prices. (See the work by John R. Hicks, especially ch. 5.) This assumption is not necessary for the proofs of all the theorems.

[13] We shall use the symbol RPT if it is valid, either for a conceptual experiment concerning the behavior of the consumer in a given price and income situation, or for actual observations on a consumer who behaves according to the basic model.

some circumstances by relating the TI to the measurable Laspeyres' price index (LI), where

$$LI = \hat{p}^1\bar{x}_0/p^0\bar{x}_0 \tag{37}$$

The following theorem may be proved:[14]

THEOREM 5: When $\overline{E}_0 \geq \overline{E}_1$, then

$$TI \leq (\overline{E}_0/\overline{E}_1)\cdot LI \tag{38}$$

The equality in (38) holds only when $\overline{E}_0 = \overline{E}_1$ and $\hat{p}^1 = p^0$. In this case Theorem 5 reduces to the result that the $TI = 1$. When $\hat{p}^1 = p^0$, but $\overline{E}_0 > \overline{E}_1$ we have a trivial upper bound on the TI; trivial because with stable prices, the $TI = 1$.

Because $\overline{E}_0 \geq \overline{E}_1$, the LI does not provide an upper bound to the TI except in the case when $\overline{E}_0 = \overline{E}_1$. When $\overline{E}_0 < \overline{E}_1$ a relationship between the TI and the LI cannot be derived.

The results of Theorems 1, 4, and 5 are combined in Table 1. The equalities of the TI to 1 and the LI in the third row of the table hold if and only if $\hat{p}^1 = p^0$, in which case $PI = LI = TI = 1$. The only situation when the PI and LI provide no information concerning the TI is when the $PI < 1$ and $\overline{E}_0 < \overline{E}_1$.

TABLE 1

PI	$\overline{E}_0<\overline{E}_1$	$\overline{E}_0=\overline{E}_1$	$\overline{E}_0>\overline{E}_1$
>1	$TI>1$	$LI>TI>1$	$(\overline{E}_0/\overline{E}_1)\cdot LI>TI>1$
$=1$	$TI\geq 1$	$LI\geq TI\geq 1$	$(\overline{E}_0/\overline{E}_1)\cdot LI>TI\geq 1$
<1	$TI \gtreqless 1$	$LI>TI$	$(\overline{E}_0/\overline{E}_1)\cdot LI>TI$

IV. Concluding Remarks

In this paper a TPI has been defined for a consumer who saves in order to maximize a simple two-period intertemporal utility function. There are several extensions of this analysis which will be dealt with in a subsequent paper. For example, we will examine how the consumer's desire for a terminal stock of savings in period 1 affects the analysis. Also the model will be extended beyond two periods and a TPI will be defined to measure price changes between different periods. In this analysis the consumer's revisions of his price expectations will play an important role.

[14] Theorem 5 is proved by showing that when $\overline{E}_0 > \overline{E}_1$, then $E_1^* < \hat{p}^1\bar{x}_0$, and when $\overline{E}_0 = \overline{E}_1$, then $E_1^* \leq \hat{p}^1\bar{x}_0$, the equality holding only when $\hat{p}^1 = p^0$.

APPENDIX

A. *Construction of Intertemporal Indifference Curves in Expenditure Space*

Assume that the level of intertemporal utility in equation (5) is held constant at some level, say W^+. The intertemporal indifference curve representing W^+ in utility space is then given by the set of utility pairs A^+ where $A^+ = [(U_0, U_1) \mid W^+ = W(U_0, U_1)]$. Now suppose that we choose any price vectors p_+^0, p_+^1, holding in periods 0 and 1, respectively. Then for each element of A^+ we may find the minimum expenditures pair (E_0, E_1) needed to attain it given the prices p_+^0 and p_+^1 holding in the two periods. When we do this for each element of the set A^+ and the chosen prices p_+^0, p_+^1, we have a set of expenditure pairs which is an intertemporal indifference curve in expenditure space representing an intertemporal utility level of W^+. We symbolize such an indifference curve by $W[p_+^0, p_+^1]$. If the level of intertemporal utility is held constant and different price vectors are chosen, the intertemporal indifference curve in expenditure space representing a level of utility of W^+ will be different in general from $W[p_+^0, p_+^1]$.

In the situations considered in this paper, the price vector of period 0 is always that of p^0, the actual prices in period 0. Hence, the intertemporal indifference curves will be symbolized by $W[p_+^1]$ where the symbol over the W indicates the level of intertemporal utility of an indifference curve and the price vector is a price vector assumed to hold in period 1. Finally, we shall assume that no two intertemporal indifference curves are identical in expenditure space unless the price situations of period 1 are the same. Thus two identical curves must represent

TABLE 2—POSSIBLE RELATIONS BETWEEN p, U, AND W, AND VALUE OF TRUE PRICE INDEX

p	U	W	TI
1. $p_e^1=p^0=\hat{p}^1$	(a) $\tilde{U}_1=\overline{U}_1=\hat{U}_1$	$\tilde{W}=\overline{W}=\hat{W}$	$TI=1$
2. $p_e^1=p^0\neq\hat{p}^1$	(a) $\tilde{U}_1=\overline{U}_1\neq\hat{U}_1$	$\tilde{W}=\overline{W}\neq\hat{W}$	$TI\neq 1$
	(b) $\tilde{U}_1=\overline{U}_1=\hat{U}_1$	$\tilde{W}=\overline{W}=\hat{W}$	$TI=1$
3. $p_e^1=\hat{p}^1\neq p^0$	(a) $\tilde{U}_1\neq\overline{U}_1=\hat{U}_1$	$\tilde{W}\neq\overline{W}=\hat{W}$	$TI\neq 1$
	(b) $\tilde{U}_1=\overline{U}_1=\hat{U}_1$	$\tilde{W}=\overline{W}=\hat{W}$	$TI=1$
4. $p_e^1\neq\hat{p}^1=p^0$	(a) $\tilde{U}_1=\hat{U}_1\neq\overline{U}_1$	$\tilde{W}=\hat{W}\neq\overline{W}$	$TI=1$
	(b) $\tilde{U}_1=\hat{U}_1=\overline{U}_1$	$\tilde{W}=\hat{W}=\overline{W}$	$TI=1$
5. $p_e^1\neq\hat{p}^1\neq p^0$	(a) $\tilde{U}_1\neq\overline{U}_1\neq\hat{U}_1$	$\tilde{W}\neq\overline{W}\neq\hat{W}$	$TI\neq 1$
	(b) $\tilde{U}_1\neq\overline{U}_1=\hat{U}_1$	$\tilde{W}\neq\overline{W}=\hat{W}$	$TI\neq 1$
	(c) $\tilde{U}_1=\overline{U}_1\neq\hat{U}_1$	$\tilde{W}=\overline{W}\neq\hat{W}$	$TI\neq 1$
	(d) $\tilde{U}_1=\hat{U}_1\neq\overline{U}_1$	$\tilde{W}=\hat{W}\neq\overline{W}$	$TI=1$
	(e) $\tilde{U}_1=\hat{U}_1=\overline{U}_1$	$\tilde{W}=\hat{W}=\overline{W}$	$TI=1$

the same level of intertemporal utility. The odd cases we are not considering are those where two different price situations in period 1 may yield identical curves (or curve segments) in expenditure space which may or may not represent the same intertemporal utility level.

B. *Alternative Relationships Between Price Vectors and Utility Levels*

For the consumer who has attained a utility of $\overline{U}_0$ in period 0 and has $\overline{E}_1$ to spend in period 1, we may imagine three possible price situations in period 1, indicated by p_e^1, $\hat{p}^1$, and p^0, the vectors of expected period 1 prices, actual period 1 prices, and period 0 prices, respectively. In each of these price situations, we may evaluate his period 1 utility levels $\overline{U}_1$, $\hat{U}_1$, or $\tilde{U}_1$, given $\overline{E}_1$, and thus his resulting intertemporal utility $\overline{W}$, $\hat{W}$, or $\tilde{W}$, given $\overline{U}_0$. However, there are five possible relationships of equality and/or inequality

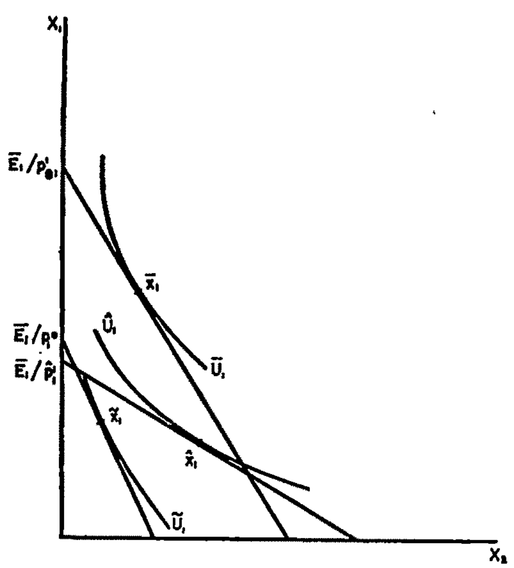


FIGURE 3

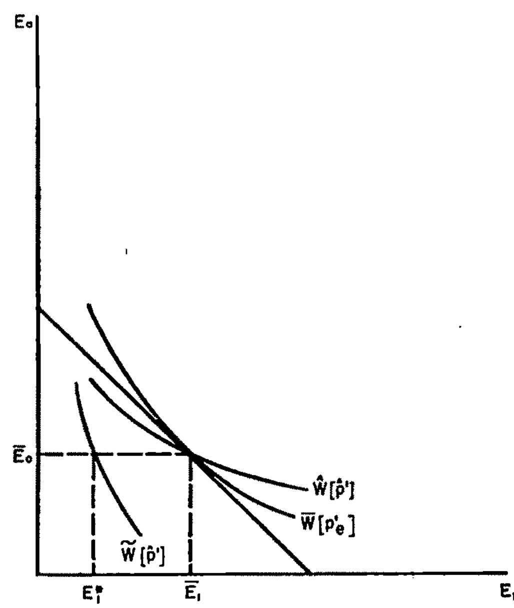


FIGURE 4

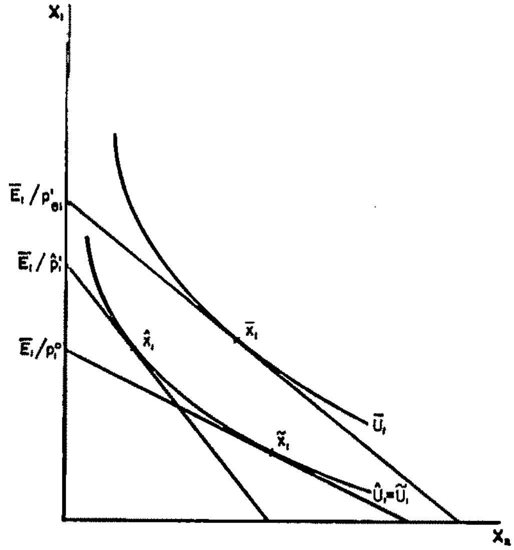


FIGURE 5

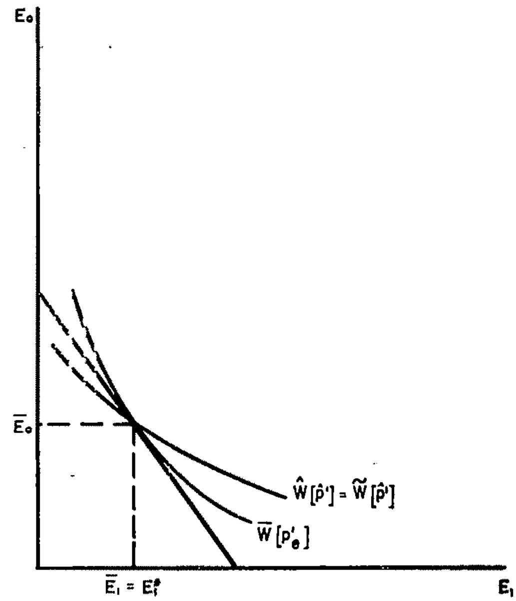


FIGURE 6

between p_e^1, $\hat{p}^1$, and p^0, for example, $p_e^1 \neq \hat{p}^1 = p^0$. For each of these relationships there will be at least one possible relationship of equality and/or inequality between $\bar{U}_1$, $\hat{U}_1$, and $\tilde{U}_1$, and equivalently between $\overline{W}$, $\hat{W}$, and $\tilde{W}$. For each of these possible sets of relationships we may determine whether or not the true price index differs from unity. All the possible situations are given in Table 2.

The *TI* is only unity when $\tilde{U}_1 = \hat{U}_1$, i.e., when $\tilde{W} = \hat{W}$, regardless of the relationship of $\bar{U}_1$ to $\tilde{U}_1$ and $\hat{U}_1$, and thus regardless of the relationship of $\overline{W}$ to $\tilde{W}$ and $\hat{W}$. Diagrammatically we may illustrate these relationships in the two-good case by using conventional indifference curve analysis, and in the *n*-good case by using intertemporal indifference curves drawn in expenditure space. As an example let us compare situation 5(a) with 5(d). In Figure 3 situation 5(a) is drawn for the case of two goods X_1 and X_2, and in Figure 4 situation 5 (a) is drawn using intertemporal indifference curves for a situation when the *TI* indicates a fall in prices. In Figures 5 and 6 situation 5(d) is drawn and in this situation the *TI* must be unity.

In all possible situations, however, the *TI* depends on a comparison between $\hat{U}_1$ and $\tilde{U}_1$, and will only be unity when $\hat{U}_1 = \tilde{U}_1$, i.e., when $\hat{W} = \tilde{W}$. From the analysis in this Appendix we see that this will certainly be the case when $p^0 = \hat{p}^1$, but this is a sufficient and not necessary condition for true price stability.

REFERENCES

F. M. Fisher and K. Shell, "Taste and Quality Change in the Pure Theory of the True Cost-of-Living Index," in J. N. Wolfe, ed., *Value, Capital, and Growth,* Edinburgh 1968.

I. Fisher, *The Theory of Interest,* New York 1930.

J. R. Hicks, *A Revision of Demand Theory,* Oxford 1956.

M. J. Ulmer, *The Economic Theory of Cost of Living Index Numbers,* New York 1949.

M. E. Yaari, "On The Consumer's Lifetime Allocation Process," *Int. Econ. Rev.,* Sept. 1964, *5,* 304–17.

Theories of Corporate Investment Behavior Revisited

By J. W. ELLIOTT*

Dale Jorgenson's investment model has been extensively tested and analyzed at aggregate and at industry levels, particularly within the manufacturing sector.[1] It has also been evaluated in this *Review* at the level of the firm.[2] In this latter evaluation, rather general conclusions were drawn from a sample of fifteen large manufacturing firms as to the performance rankings of alternative investment models. These rankings were based principally on the size of residual variation resulting from successively inserting each model into the flexible accelerator mechanism fit to time-series data of the individual firms. Jorgensen and Siebert interpreted the results as showing the superiority of the neoclassical theory of investment over other considered alternatives.

Even aside from lingering criticisms of Jorgenson's model such as that of Robert Eisner (1970), and Eisner and M. Ishaq Nadiri, the Jorgenson and Siebert (JS) firm-level results are based on a rather meager sample (fifteen firms) given the apparently general objectives of their work.

Greater meaning may be attached to the JS conclusions either by expanding their time-series sample to cover more firms or to conduct their comparison of alternative investment models on a cross-sectional basis. The cross-sectional alternative is attractive in that a great many cells of firm-level data may be evaluated in a manner leading to results with some general significance. Patterns which emerge from cross-sectional tests on individual firm data will reflect relationships persisting across firm and industry boundaries.

Further, as Edwin Kuh has pointed out, the impact of "inter-firm dynamic specification difficulties will be less in cross-sections since market or cyclical factors faced in common by all firms in an industry will not affect inter-firm variability and therefore will exert a minor influence on cross-section estimates" (p. 185). However, as Kuh also recognizes, cross-sections are likely to be relatively poor vehicles for analysis when dynamic disturbances are present and unaccounted for by the model (p. 328). This implies that the reasonableness of cross-section results depends somewhat on the correctness of the dynamic model specification. In the present case, this amounts to accepting the flexible accelerator and Jorgenson's generalized Pascal distributed lag structure a priori.

In addition, the use of time-series data to empirically fit the lag structure in the JS study led to measured coefficients representing a single characterization of the lag pattern over the sample period for each firm. Following the same procedure in cross-sections leads to selection of a single measured lag structure for each year best representative of the sample firms. To the extent that parameters of the true lag pattern vary intertemporally there may be advantages in the cross-section approach which outweigh disadvantages associated with selecting a single lag structure for the entire sample.

In the JS study it was simply concluded a priori that "In order to specify the lag structure correctly we concentrate on time series data for individual firms" (1968a, p. 690). When the objective is to extract generalizable patterns from firm-level data, the choice of time-series does not seem as straightforward as this statement would have us believe. Indeed, it is difficult to define a clear-cut case a priori for the direction of analysis of firm-level data in such a case.

* University of Wisconsin, Milwaukee. I wish to thank Dale Jorgenson and Calvin Siebert for supplying representative data on their variables, Naime Sherbiny and several unnamed referees for their constructive help in this project, and the Graduate School of the University of Wisconsin-Milwaukee for computer facilities and assistance.

[1] See for example, Jorgenson (1963) and Jorgenson and Robert Hall, plus several of Jorgenson's more recent works.

[2] See Jorgenson and Siebert (1968a). The identical sample is used in Jorgenson and Siebert (1968b).

TABLE 1

Industry	Number of Firms	Industry	Number of Firms
Miscellaneous Metals	4	Steel	5
Lead and Zinc	3	Foundries—Gray Iron	1
Gold Mining	1	Copper	2
Coal—Bituminous	1	Aluminum	3
Oil—Crude Producers	1	Machinery—Metal Fabricating	2
Sulphur	1	Building Materials—Heating, Air Conditioning and Plumbing	4
Food—Packaged Foods	3	Metal Work—Miscellaneous	3
Food—Dairy Products	3	Machinery—Steam Generating	1
Food—Canned Food	3	Machinery—Agricultural	3
Food—Prepared Foods for Animals and Fowl	1	Machinery—Construction and Materials Handling	4
Food—Bread and Cake Bakers	4	Machinery—Oil Well	2
Food—Biscuit Bakers	2	Machine Tools	5
Sugar—Cane Refiners	1	Machinery—Specialty	5
Sugar—Beet Refiners	1	Machinery—Industrial	2
Beverages—Brewers	3	Machinery—General Industrial	1
Beverages—Distillers	5	Electrical and Electronic Leaders	4
Vegetable Oil	1	Electrical Equipment	4
Tobacco—Cigarette Manufacturing	4	Electrical Instruments and Controls	2
Tobacco—Cigar Manufacturing	1	Radio-T.V. Manufacturers	1
Textile Products	2	Automobiles	2
Textile—Apparel Manufacturing	2	Auto Trucks	1
Home Furnishings	3	Auto Parts and Accessories	4
Paper	5	Aerospace	3
Paper—Containers	1	Aircraft—General	1
Publishing	2	Railroad Equipment	2
Chemicals	5	Engineering Lab. and Research Equipment	2
Drugs	5	Optical	1
Soaps	1	Watches	2
Cosmetics	1	Manufacturing Industries	1
Chemicals and Chemical Preparations	1	Air Transport	4
Oil—Integrated Domestic	5	Telephone Companies	1
Oil—Integrated International	4	Wholesale—Other	1
Building Materials—Roofing and Wallboard	5	Retail—Department Stores	2
Tire and Rubber Goods	5	Retail—Variety Stores	2
Shoes	1	Retail—Food Chains	3
Containers—Metal and Glass	3		
Building Materials—Cement	4	Total	184

In expanding the basis for the JS comparison, this led to analysis of a much enlarged firm-level sample on both a time-series and cross-sectional basis, which taken together should help clarify the meaning of resulting conclusions.

Accordingly, data on the JS variables were developed from 184 firms inside and outside manufacturing. These firms are drawn from 72 four-digit SIC-level industries. The specific industries and numbers of firms are shown in Table 1. Data on firm variables were obtained for 1953–67 from Standard and Poor's Compustat data bank. Selection of firms was made on the basis of: a) available data; b) no greater than five firms per industry; c) compatibility in accounting practice over the sample period. These criteria were aimed at developing a broadly based sample of large firms.

Analysis of the data was undertaken for each of the years 1953–67, and for each of the 184 firms separately. One definitional change was made from the JS definitions, relating to

deflators. Because the cross-sectional test invites heteroscedasticity due to inter-firm differences, all variables employed in the JS analysis which were not unit free were deflated by the firm's total assets, thus greatly reducing inter-firm differences in magnitude. This made our investment measure a representation of capital investment intensity rather than direct volume of spending.[3]

This is not a trivial difference and raises the possibility that our enlarged results may only be applicable to intensity of investment (per dollar of total assets) and not to raw volume of investment. Fortunately, the smoothness of typical total asset series compared to investment series appears to reduce this possibility greatly. In addition, it is difficult to believe that fundamental inter-model differences in explanatory power will not emerge given investment intensity rather than volume as the dependent variable.

To compare major alternative investment models, JS employed the well-known flexible accelerator framework in which:

$$(1)\quad K_t - K_{t-1} = (1-\lambda)(K_t^* - K_{t-1})$$

where K_t and K_t^* are actual and desired levels of capital in period t. Assuming replacement to be a constant proportion, δ, of capital enables the change in capital to be expressed as:

$$(2)\quad K_t - K_{t-1} = E_t - \delta K_{t-1}$$

where E_t is the level of new capital expenditures. Expressions (1) and (2) may be combined to produce a flexible accelerator framework for analysis of investment expenditures:

$$(3)\quad E_t = (1-\lambda)(K_t^* - K_{t-1}) + \delta K_{t-1}$$

To complete the model it is necessary to specify the determinants of the desired level of capital stock. It is at this point that theoretical conflicts most clearly occur, as different investment theories have emphasized the primacy of varying principal influences on K_t^*. Based on this view, the JS comparison defined alternative representations of K^* reflecting major branches of investment theory. The statistical importance of each K^* measure was evaluated by its separate insertion in an operational version of the flexible accelerator framework of (3).

This operational version derives from generalizing the flexible accelerator and assuming the distributed lag influence of each K^* on E_t follows a generalized Pascal distribution. JS show that generalizing (3) gives

$$(4)\quad E_t = \sum_{r=0}^{\infty} \mu_r (K_{t-r}^* - K_{t-r-1}^*) + \delta K_{t-1}$$

where the μ_r are weights which are non-negative and sum to unity. Incorporating their lag specification gives:

$$(5)\quad \begin{aligned} E_t = &\sum_{r=0}^{2} \mu_r (K_{t-r}^* - K_{t-r-1}^*) \\ &- \sum_{i=1}^{2} w_i (E_{t-i} - \delta K_{t-i-1}) + \delta K_{t-1} \end{aligned}$$

which is the specific model used to compare the relevance of alternative desired capital stock specifications. In this model, up to three desired capital stock and up to two lagged net investment terms were admitted into the specification of the distributed lag, based on their explanatory impact on investment. Terms were added to the equation if they lowered the standard error.

Four theoretical alternatives are reflected in the competing specifications of the K^* terms in (5). These include:

(I_{cl}) The neoclassical formulation of Jorgenson which sees K^* as proportional to output (Q) and the user cost of capital services (C), i.e.,

$$K^* = \propto \left(\frac{pQ}{C}\right)$$

where p = product price and $\propto$ is a proportionality constant.

[3] The National Industrial Conference Board fixed capital stock deflators were used at several points in the analysis, following Jorgenson. However, in our case involving some nonmanufacturing firms in the sample, their use relies on the assumptions that price changes for capital goods purchased by nonmanufacturing firms were approximately the same as price increases on purchases by manufacturing firms.

(I_a) A capacity utilization or accelerator model is based on output changes devised to represent the work of Kuh, Eisner, and Hickman:

$$K^* = \propto Q$$

(I_l) A model measuring cash flow, (L), proposed to measure cash flow effects similar to those found important by Grunfeld and Kuh:

$$K^* = \propto L$$

(I_{exp}) An expected profits model built on the market value of the firm, (V) as suggested by Grunfeld:

$$K^* = \propto V$$

For details of the statistical formulation of these competing investment models, the reader should consult the statistical appendix of a working paper available from JS.

JS also considered a second version of the neoclassical model which treated capital gains effects explicitly. This second version is considerably more demanding computationally; therefore, it did not seem necessary to include both versions in the analysis. Moreover, the two versions ranked next to each other in all performance comparisons.

Following JS, current, one-, and two-lagged changes in each alternative K^* representation (I variable) were inserted into (5) and retained if they contributed to the explanation of investment intensity by lowering the standard error of estimate. One- and two-lagged values of the net investment term of (5) were included on the same basis. With this procedure, the most empirically significant I model should provide a superior explanation of fluctuations in investment intensity in a high number of sample firms. Accordingly, the four competing I models are ranked on the basis of the size of the standard error resulting from their employment as explanatory variables for investment intensity in (5).

Some representative results of the time-series tests are shown in Table 2. Coefficients, standard errors, and goodness-of-fit data are shown for eleven firms, drawn from the eleven major industrial groupings in the sample. These groupings are discussed below in connection with the homogeneity tests. The firms shown are the first in each major industrial grouping. Similar data on the cross-sectional tests are given in Table 3 for each of the 15 years in the sample. The time-series R^2 values are comparable with those obtained by JS, thus suggesting that the use of investment intensity in place of volume did not noticeably damage the level of explanation. However, for the expanded time-series firms, the measured direction of effect of the desired capital stock proxies is not consistent across firms. This contrasts with the JS results which were uniform in direction of effect for all desired capital stock proxies except the liquidity model.[4]

The sign inconsistencies observed in the present work may be simply due to the multicollinearity frequently found. In fact, in many instances, attempts to even compute the determinant of the correlation matrix of independent variables in order to calculate a chi-square statistic for Wilks-Bartlett's approximate test for multicollinearity were frustrated by the severity of the condition. In any event, the frequency with which multicollinearity was evident in the present results reduces greatly the inferences we may draw from observed values of individual coefficients. Fortunately, this condition does not damage the usefulness of the present results for comparing alternative desired capital stock proxies on the basis of their impact on explained variation in investment intensity.

In the cross-sections, the level of explanation remains reasonably high for most desired capital stock proxies in most years.[5] This as much as anything suggests broadly based intertemporal support for the flexible accelerator mechanism. The inconsistency in the direction of effect of all the desired capital stock proxies observed in the time-series data remains, as does the appearance of multicollinearity.

[4] Although coefficients were not included in the JS article, they were included in a longer working paper supplied by the authors.

[5] For the degrees of freedom involved ($n=184$), R^2 values of as low as .10 are significantly different from zero beyond the 1 percent level.

TABLE 2—REPRESENTATIVE TIME-SERIES RESULTS

	Desired Capital Stock			Net Investment						
	t	t−1	t−2	t−1	t−2	K_{-1}	Constant	R^2	Standard Error	*DW*
				1. American Metal Climax						
I_{cl}			−.0051	.7354		−.0795	.0693	.468	.0297	2.2
			1.33	*1.76*		*0.21*	*0.82*			
I_{exp}	−.0554	−.0085	−.0678	1.354		−.5914	.1836	.602	.0284	2.23
	1.30	*0.18*	*1.41*	*2.67*		*1.22*	*1.64*			
I_a				.6986		.0399	.0408	.384	.0307	2.32
				1.63		*0.11*	*0.48*			
I_l				.6986		.0399	.0408	.384	.0307	2.32
				1.63		*0.11*	*0.48*			
				2. Consolidated Foods						
I_{cl}	−.0069				−.3301	.2093	.0037	.485	.0234	1.49
	1.90				*1.48*	*1.90*	*0.08*			
I_{exp}	−.0779	−.0544	−.0730		−.3986	.3045	−.0208	.550	.0242	2.73
	1.68	*1.19*	*1.63*		*1.57*	*2.31*	*0.37*			
I_a	−.0459	.0281				.1208	.0310	.481	.0235	1.07
	2.23	*1.32*				*1.08*	*0.65*			
I_l	4.196	2.052			−1.113	.1563	.0286	.572	.0224	3.13
	2.32	*1.87*			*2.92*	*1.50*	*0.64*			
				3. American Tobacco						
I_{cl}	.0012	.0013	.0014	.0137	−.5124	.4155	−.0329	.892	.0038	2.67
	1.98	*2.10*	*2.11*	*0.04*	*1.52*	*3.49*	*2.78*			
I_{exp}	−.0051	.0101	−.0086	1.046	−1.049	.2664	−.0196	.948	.0026	1.74
	1.03	*1.44*	*1.75*	*2.48*	*2.19*	*3.56*	*2.78*			
I_a		−.0517		.6754		.1677	−.0105	.935	.0025	1.95
		4.60		*3.27*		*3.38*	*2.22*			
I_l				.3600	−.3751	.2776	−.0195	.826	.0041	2.44
				1.15	*1.05*	*2.90*	*2.15*			
				4. Crown Zellerbach						
I_{cl}	.0436					.3162	−.0735	.466	.0251	1.44
	3.17					*2.01*	*0.89*			
I_{exp}	.1060	−.0421		1.222	−.6603	.0757	.0291	.650	.0234	1.90
	2.42	*0.99*		*3.49*	*2.05*	*0.36*	*0.27*			
I_a				.5605		−.1108	.1308	.300	.0287	1.75
				2.19		*0.60*	*1.38*			
I_l		1.257	.6664	.5514		−.1016	.1291	.474	.0272	2.19
		1.25	*0.71*	*1.71*		*0.47*	*1.21*			
				5. Ashland Oil						
I_{cl}	−.0611				−.2418	−.4086	.3691	.516	.0316	2.72
	2.54				*1.46*	*2.73*	*4.07*			
I_{exp}	−.1331				−.1833	−.3816	.3548	.470	.0331	1.90
	2.22				*1.08*	*2.48*	*3.78*			
I_a			.0953	−.2752		−.2601	.2852	.375	.0359	2.29
			1.18	*1.30*		*1.56*	*2.79*			
I_l				−.2768		−.3233	.3205	.296	.0365	2.42
				1.29		*2.01*	*3.23*			
				6. National Gypsum						
I_{cl}					−.2585	−.3889	.1666	.419	.0327	1.80
					1.42	*2.54*	*4.12*			
I_{exp}	.0528				−.2839	−.3301	.1495	.470	.0326	1.98
	1.03				*1.55*	*2.02*	*3.43*			
I_a	−.0497	−.1117	−.0353		−.2735	−.2197	.1165	.629	.0301	2.42
	1.39	*2.06*	*1.11*		*1.59*	*1.36*	*2.68*			
I_l	−1.373	−1.788	−.6097		.5386	−.0960	.0842	.878	.0173	1.97
	5.17	*5.70*	*3.60*		*3.03*	*0.87*	*2.89*			

Note: Numbers in *italics* below each coefficient are t-values.

TABLE 2—(*Continued*)

	Desired Capital Stock			Net Investment						
	t	t−1	t−2	t−1	t−2	K_{-1}	Constant	R^2	Standard Error	*DW*
				7. Armco Steel						
I_{cl}				.4935	−.7527	−.0062	.0872	.519	.0239	1.44
				2.57	*2.63*	*0.07*	*2.24*			
I_{exp}				.4935	−.7527	−.0062	.0872	.519	.0239	1.44
				2.57	*2.63*	*0.07*	*2.24*			
I_a		.1510	.0941	.5561	−.4512	−.0666	.1113	.717	.0203	1.80
		2.23	*1.27*	*2.62*	*1.59*	*0.86*	*3.20*			
I_l			.9625	.4127	−.2832	.0100	.0772	.753	.0179	1.98
			3.09	*2.81*	*1.08*	*0.15*	*2.62*			
				8. Combustion Engineering						
I_{cl}	.0405	.1032		.0776	−1.156	−.0523	.0681	.779	.0203	2.19
	2.24	*4.58*		*0.28*	*4.30*	*0.27*	*1.58*			
I_{exp}		.1339	.1431	.4383		−.4855	.1536	.394	.0319	2.35
		1.64	*1.37*	*1.49*		*1.50*	*2.24*			
I_a	.1982	.1167			−.9975	.0140	.0481	.513	.0286	1.76
	3.21	*1.70*			*2.39*	*0.96*	*0.85*			
I_l		−2.906	1.099	.9687		−.5433	.1707	.528	.0282	2.64
		2.60	*1.17*	*3.12*		*1.92*	*2.63*			
				9. Honeywell						
I_{cl}	.0739		−.0237	−.6803	−1.124	1.210	−.1885	.844	.0160	2.77
	5.29		*3.87*	*2.46*	*4.50*	*4.33*	*2.77*			
I_{exp}	.0222	−.0210			−.6491	.8209	−.1255	.476	.0278	2.93
	1.07	*0.99*			*1.65*	*2.71*	*1.57*			
I_a	−.0785		.2323	−.4382		.6547	−.0737	.729	.0200	2.41
	1.65		*3.51*	*1.37*		*2.86*	*1.27*			
I_l			.6942			.3142	.0027	.457	.0259	2.15
			1.98			*1.82*	*0.05*			
				10. Chrysler						
I_{cl}	.0122			.3092		−.1913	.1388	.724	.0216	2.24
	2.43			*2.36*		*3.42*	*3.84*			
I_{exp}	.0877		−.0571	.3280	.2773	−.2024	.1400	.867	.0166	2.23
	3.60		*1.56*	*2.53*	*2.00*	*3.94*	*4.52*			
I_a	.0167			.4017		−.2282	.1564	.643	.0246	1.73
	1.44			*2.49*		*3.62*	*3.95*			
I_l				.3107		−.2717	.1750	.575	.0257	1.60
				2.00		*3.19*	*4.48*			
				11. American Airlines						
I_{cl}	−.0992	−.0818	−.0451			−.3916	.4270	.346	.0542	1.55
	2.14	*1.84*	*1.00*			*1.68*	*1.77*			
I_{exp}	−.2560		−.2123	.3603		−.4010	.4252	.354	.0539	2.01
	1.66		*1.88*	*1.35*		*0.92*	*1.52*			
I_a	−.4237					−.1404	.2647	.415	.0468	0.98
	2.92					*0.54*	*1.53*			
I_l	−1.967	1.275				−.0427	.2090	.302	.0534	2.40
	1.93	*1.24*				*0.14*	*1.01*			

These cross-sectional results relative to the I_a variable present an opportunity for some comparisons with recent empirical work by Robert Eisner. In 1969, he explored the cross-sectional relevance of a number of explanatory mechanisms involving sales change, profits, and depreciation in explaining variation in a measure of investment intensity very similar to that used in the present work. Among his results is the finding that the distributed lag effect of his sales-change variable is unstable intertemporally in cross-sectional tests. Eisner suggests this may be due to marked differences in growth rates in sales before and after 1960. The I_a variable of the present comparison is roughly com-

TABLE 3—CROSS-SECTIONAL RESULTS
1953–67

	Desired Capital Stock			Net Investment						
	t	t−1	t−2	t−1	t−2	K_{-1}	Constant	R^2	Standard Error	*DW*
					1953					
I_{cl}	.0027	−.0033	.0037	.3499	−.0447	.1133	.0152	.451	.0437	2.09
	0.96	*1.18*	*1.33*	*4.69*	*0.61*	*7.13*	*1.98*			
I_{exp}	.0719	.0275	.0504	.2795	−.0729	.1126	.0120	.501	.0417	2.07
	3.29	*1.87*	*2.91*	*3.72*	*1.04*	*7.48*	*1.70*			
I_a				.3370		.1132	.0129	.433	.0439	2.08
				4.60		*7.31*	*1.74*			
I_l	−.3558	.3023		.3825	−.0793	.1162	.0155	.465	.0431	1.99
	1.67	*1.97*		*5.15*	*1.11*	*7.56*	*2.12*			
					1954					
I_{cl}	−.0027		.0052	.5318	−.2089	.1308	.0208	.404	.0432	1.64
	0.99		*2.81*	*5.24*	*2.18*	*8.06*	*2.68*			
I_{exp}	.0162		.0118	.5163	−.2233	.1443	.0094	.388	.0438	1.70
	1.65		*0.95*	*5.01*	*2.28*	*8.47*	*1.15*			
I_a	−.0940	−.0369		.5052	−.2336	.1319	.0132	.429	.0423	1.58
	4.07	*1.40*		*4.99*	*2.48*	*8.36*	*1.71*			
I_l	−.5463			.5296	−.1813	.1355	.0081	.416	.0426	1.67
	3.61			*5.30*	*1.91*	*8.64*	*1.09*			
					1955					
I_{cl}	.0016	.0009		.4169		.1244	.0026	.524	.0402	2.13
	1.50	*1.23*		*6.14*		*9.49*	*0.38*			
I_{exp}	.0148	.0248	.0548	.4413		.1183	.0008	.553	.0390	2.09
	0.94	*1.20*	*3.59*	*6.60*		*9.12*	*0.13*			
I_a				.4134		.1254	.0004	.517	.0402	2.13
				6.09		*9.63*	*0.06*			
I_l	.0867		.1940	.4073		.1271	.0010	.524	.0402	2.14
	0.71		*1.49*	*5.96*		*9.73*	*0.16*			
					1956					
I_{cl}	−.0014	.0017		.1693		.1346	.0249	.397	.0482	1.66
	1.15	*0.94*		*2.88*		*8.47*	*3.18*			
I_{exp}	−.0227	.0520	.0378	.1729	−.0800	.1328	.0197	.430	.0472	1.74
	1.49	*3.08*	*2.11*	*2.94*	*0.88*	*7.82*	*2.35*			
I_a	−.0589	−.0550	−.0444	.1275	−.0733	.1480	.0173	.429	.0472	1.83
	2.64	*2.31*	*2.30*	*2.15*	*0.82*	*8.87*	*2.05*			
I_l	−.3097			.1440		.1345	.0240	.399	.0480	1.74
	1.73			*2.43*		*8.51*	*3.11*			
					1957					
I_{cl}	.0032		.0010	.2004	−.0794	.0946	.0219	.303	.0383	1.90
	1.92		*1.03*	*2.89*	*1.19*	*6.86*	*3.31*			
I_{exp}	.0077			.1989	−.0775	.0951	.0202	.288	.0386	1.89
	0.72			*2.84*	*1.15*	*6.73*	*3.13*			
I_a	−.0248	.0491	−.0288	.2452	−.1094	.0922	.0188	.342	.0373	1.90
	1.56	*3.01*	*1.76*	*3.52*	*1.64*	*6.84*	*2.86*			
I_l	−.1690	.3230	−.1771	.2352	−.0911	.0935	.0187	.340	.0374	1.92
	1.24	*2.54*	*1.18*	*3.39*	*1.39*	*6.78*	*2.94*			
					1958					
I_{cl}	−.0031	−.0031		.2910		.1303	.0174	.435	.0434	1.81
	2.04	*1.83*		*4.12*		*8.65*	*2.33*			
I_{exp}	.0172		.0228	.2688		.1301	.0098	.435	.0434	1.95
	1.29		*1.52*	*3.76*		*8.57*	*1.36*			
I_a	−.0499	.0153	−.0172	.2938		.1339	.0119	.451	.0429	1.86
	2.70	*0.75*	*0.92*	*4.10*		*9.00*	*1.64*			
I_l	−.3360	−.2676	−.3797	.2958	−.0389	.1270	.0175	.461	.0426	1.89
	2.41	*1.56*	*3.03*	*4.21*	*0.74*	*8.19*	*2.39*			

Note: Numbers in *italics* below each coefficient are t-values.

TABLE 3—(*Continued*)

	Desired Capital Stock			Net Investment					Standard	
	t	t−1	t−2	t−1	t−2	K_{-1}	Constant	R^2	Error	DW
					1959					
I_{el}	.0053	.0024		.3725	−.0422	.1079	.0138	.371	.0373	1.68
	2.00	*1.52*		*5.13*	*0.63*	*7.80*	*2.10*			
I_{exp}				.3714	−.0392	.1054	.0122	.351	.0376	1.69
				5.06	*0.58*	*7.57*	*1.88*			
I_a	−.0354	.0158	.0137	.3365		.0982	.0194	.372	.0372	1.72
	1.70	*0.98*	*0.82*	*4.69*		*7.47*	*2.78*			
I_l	−.4099		.0606	.3412		.1012	.0172	.378	.0369	1.65
	2.62		*0.52*	*4.87*		*7.79*	*2.61*			
					1960					
I_{el}	.0019			.4200		.1261	.0015	.515	.0406	2.12
	1.59			*6.08*		*9.59*	*0.23*			
I_{exp}	.0172	.0265	.0559	.4418		.1175	.0019	.545	.0395	2.08
	1.08	*1.27*	*3.60*	*6.47*		*8.95*	*0.30*			
I_a		.0163	.0235	.4355		.1242	.0024	.514	.0407	2.16
		0.89	*1.43*	*6.14*		*9.38*	*0.35*			
I_l	.1691	.3221	.4125	.4183		.1239	.0026	.534	.0400	2.14
	1.34	*1.58*	*2.89*	*6.02*		*9.48*	*0.40*			
					1961					
I_{el}			.0024	.2278	.0395	.0874	.0180	.347	.0323	1.95
			1.06	*3.66*	*0.55*	*7.01*	*3.26*			
I_{exp}	−.0050			.2459		.0867	.0176	.343	.0323	1.96
	0.72			*4.31*		*6.98*	*3.18*			
I_a				.2398		.0871	.0180	.341	.0323	1.95
				4.25		*7.02*	*3.27*			
I_l	−.3195	−.1949		.2088	.0694	.0879	.0153	.368	.0319	1.90
	2.63	*1.53*		*3.30*	*0.97*	*7.12*	*2.70*			
					1962					
I_{el}		−.0011	−.0028	.3152	.0654	.0846	.0237	.295	.0343	2.02
		0.84	*1.24*	*3.14*	*1.02*	*5.09*	*3.90*			
I_{exp}		−.0079	−.0146	.3424	.0961	.0797	.0256	.309	.0340	2.05
		1.09	*1.92*	*3.42*	*1.48*	*5.63*	*4.07*			
I_a	−.0680	−.0124	−.0240	.2835	.0655	.0867	.0230	.338	.0333	2.03
	3.64	*0.67*	*1.39*	*2.89*	*1.01*	*5.38*	*3.78*			
I_l	−.2239			.2954	.0736	0870	.0230	.302	.0340	2.04
	1.88			*2.94*	*1.16*	*5.40*	*3.95*			
					1963					
I_{el}	−.0015		.0010	.2468	.1460	0783	.0241	.274	.0322	1.87
	1.68		*0.74*	*3.22*	*1.60*	*6.08*	*4.34*			
I_{exp}			−.0062	.2471	.1546	.0778	.0238	.266	.0323	1.85
			0.90	*3.23*	*1.68*	*6.03*	*4.25*			
I_a	−.0392	.0232		.2824	.1238	.0797	.0226	.284	.0320	1.87
	1.87	*1.30*		*3.59*	*1.35*	*6.22*	*4.05*			
I_l	−.1393			.2582	.1414	.0783		.266	.0323	1.87
	0.95			*3.29*	*1.53*	*6 03*				
					1964					
I_{el}		−.0008		.3035		.1361	.0138	.366	.0358	1.78
		0.89		*3.39*		*9.32—*	*2.23*			
I_{exp}	.0168	.0045		.3077		.1362	.0136	.371	.0358	1.78
	1.41	*0.60*		*3.42*		*9.25*	*2.00*			
I_a	−.0339	.0418		.3261		.1334	.0156	.382	.0354	1.85
	1.65	*1.76*		*3.59*		*9.22*	*2.53*			
I_l	−.1451	.4060	.3271	.3516		.1341	.0132	.399	.0350	1.85
	0.83	*2.30*	*2.33*	*4.03*		*9.36*	*2.12*			

TABLE 3—(*Continued*)

	Desired Capital Stock			Net Investment						
	t	t−1	t−2	t−1	t−2	K_{-1}	Constant	R^2	Standard Error	DW
					1965					
I_{cl}	−.0069	−.0032	−.0037	.3360		.1143	.0280	.293	.0426	1.81
	1.65	*1.68*	*2.06*	*3.35*		*5.98*	*3.75*			
I_{exp}	.0096			.3501		.1083	.0304	.276	.0429	1.83
	0.77			*3.50*		*5.68*	*4.08*			
I_a	−.0605		−.0244	.3216		.1146	.0290	.300	.0423	1.83
	2.54		*0.88*	*3.24*		*6.05*	*3.95*			
I_t	−.2183	.1864	.2716	.3377		.1083	.0300	.291	.0427	1.85
	1.38	*0.84*	*1.35*	*3.35*		*5.69*	*3.84*			
					1966					
I_{cl}	−.0112	.0038		.3616	.0912	.1240	.0321	.369	.0434	2.16
	2.55	*1.59*		*4.29*	*0.87*	*5.72*	*3.99*			
I_{exp}	.0162	.0351	.0256	.3330		.1394	.0232	.375	.0432	2.03
	1.14	*2.65*	*1.79*	*3.99*		*6.67*	*2.82*			
I_a	−.0417			.3677		.1337	.0279	.352	.0437	2.11
	1.86			*4.45*		*6.50*	*3.57*			
I_t	−.8097	.4848	.2069	.4293		.1299	.0241	.409	.0420	2.07
	3.55	*3.09*	*1.12*	*5.30*		*6.52*	*3.07*			
					1967					
I_{cl}		−.0038		.3596	−.1109	.1117	.0262	.436	.0327	1.78
		1.21		*5.82*	*1.70*	*6.58*	*4.30*			
I_{exp}		−.0103	−.0123	.3823	−.1147	.1130	.0248	.438	.0328	1.81
		0.95	*1.22*	*6.03*	*1.75*	*6.47*	*3.98*			
I_a	−.0338	−.0435		.3290	−.1106	.1157	.0234	.462	.0320	1.74
	1.98	*2.64*		*5.36*	*1.74*	*7.01*	*3.94*			
I_t	−.3981	−.3950		.3556	−.0901	.1091	.0231	.468	.0319	1.74
	2.98	*2.25*		*5.86*	*1.41*	*6.63*	*3.92*			

Note: Numbers in *italics* below each coefficient are t-values.

parable with his sales change variable.[6] Our sign inconsistencies in the I_a variable in the present work are generally similar to his results, but do not show any particular breakpoint in pattern before and after 1960. However, at least in the present analysis, the observed inconsistencies may be entirely due to a suspiciously high level of multicollinearity.[7]

Troublesome as these sign inconsistencies might be, they do not directly relate to the principal data upon which the JS conclusions are based. These data emerge for each firm from a ranking of standard errors resulting from the best-fitting distributed-lag specification for each desired-capital-stock proxy. For our expanded sample, such residual-variation comparisons have been prepared for both the time-series and cross-sectional results, and shown in Table 4. In Table 4, the standard error for each of the four models has been ranked for each of the 184 firms in the enlarged time-series sample. The model listed vertically on the Y axis had a lower standard error than the model listed horizontally on the X axis in the number of cases shown.

One aspect of these expanded time-series rankings of particular interest is the extent to which they reflect systematic industry differences in model rankings. If important differences in residual variance rankings is

[6] Eisner deflated his sales change terms by a moving average of past or base period sales change, while our I_a sales change term was deflated by past total assets. This prohibits comparisons between coefficient magnitudes. However, the significance of t-values can be reasonably examined.

[7] However, Wilks-type F-values suggest that the presence of multicollinearity in the present cross-section is most severe among desired capital stock terms. This indicates the total summed effect of a desired capital stock coefficient will be more reliable than individual terms. However, substantial intertemporal instability continues to be present in the summed cross-sectional effect of the I_a term, which is suggestive of Eisner's result.

TABLE 4—MODEL RANKINGS[a]

JS Comparisons: 15 Firms (1949–63)
Time-Series

	I_{cl}	I_{exp}	I_a	I_l
I_{cl}	—	9	11	10
I_{exp}	5	—	9	10
I_a	4	6	—	10
I_l	4	4	5	—

Enlarged Comparisons: 184 Firms (1953–67)
Time-Series

	I_l	I_a	I_{cl}	I_{exp}
I_l	—	88	94	95
I_a	86	—	88	102
I_{cl}	81	88	—	97
I_{exp}	79	75	74	—

Enlarged Comparisons: 15 Years (1953–67); 184 Firms
Cross-Section

	I_l	I_a	I_{exp}	I_{cl}
I_l	—	9	9	13
I_a	6	—	9	9
I_{exp}	4	4	—	7
I_{cl}	2	3	5	—

[a] The overall JS ranking is: I_{cl}; I_{exp}; I_a; I_l. The neoclassical I model (containing capital gains effects) has been eliminated from this comparison. It ranked slightly above the simpler neoclassical II model, I_{cl} (superior in 8 of 15 firms in the standard error criterion) and was almost identical on other criteria used.

attributable to industry, then our overall rankings become a product of the choice of industrial sample, and the meaning of the results of Tables 4 and 5 is altered accordingly. To test for industry-homogeneity of the obtained rankings, a chi-square test was employed. Combinations of two-digit level industries were formed to produce groupings of at least ten firms, having either single or side-by-side two-digit codes. The overall ranking proportions for each model were used as the basis of the expected frequencies in each industrial grouping. A smallest-cell overall proportion of .216 required at least ten firm groupings in order to produce a minimum expected cell frequency greater than 2, a requirement for the test. The industrial groupings, numbers of firms, and calculated chi-square values are shown in Table 6. As seen, only one industrial segment, Food and Beverages was significantly nonhomogeneous at the 5 percent level, producing a value significant at approximately the 4 percent level. This suggests the obtained results are largely free of substantial industry effects, and are not particularly due to the industry composition of the enlarged sample.

However, since this may not be true of the Food and Beverage segment of our sample, it is of interest to examine these results separately. Table 7 shows model comparisons similar to those of Table 4 for the twenty-seven Food and Beverage firms in the sample. As seen, the impact of the nonhomogeneity is mainly upon the ranking of the neoclassical model, which moves from an overall ranking of third to first among these firms, while other models remain in the same relative position as in the overall rankings. This suggests that the appearance of twenty-seven Food and Beverage firms in the sample may somewhat unduly influence the results in favor of the neoclassical explanation of investment. In Table 8, the time-series comparisons of Table 4 are recomputed omitting the Food and Beverage firms. As seen, little impact on the overall rankings results. Taken together, the results of Tables 6, 7, and 8 suggest that our rankings and model comparisons are largely unaffected by the industry composition of the sample, although some

TABLE 5—SIGNIFICANT COEFFICIENTS FOR DESIRED CAPITAL STOCK

	t	t−1	t−2	Total
JS Results				
I_{cl}	14	10	4	28
I_{exp}	13	7	3	23
I_a	9	9	1	19
I_l	4	8	2	14
Enlarged Time-Series				
I_a	88	86	93	267
I_l	93	83	82	258
I_{cl}	93	73	84	250
I_{exp}	94	82	74	250
Cross-Section				
I_l	15	9	8	32
I_{exp}	11	8	10	29
I_a	11	10	7	28
I_{cl}	11	10	7	28

TABLE 6—HOMOGENEITY TEST OF RANKINGS

Group	Two-Digit Compustat Code	Industry Type	Number of Firms	Chi-Square
1	10, 12, 13, 14	Metals and Mining	11	15.570
2	20	Food and Beverages	27	17.760[a]
3	21, 22, 23, 25	Tobacco, Textile, Furnishings	12	13.462
4	26, 27, 28	Paper, Chemicals, Drugs	21	10.179
5	29	Oil, Building Materials	14	9.772
6	30, 31, 32	Rubber Products, Containers, Cement	13	16.669
7	33, 34	Steel, Aluminum, Copper, etc.	20	8.736
8	35	Machinery	23	5.280
9	36	Electrical Equipment	11	5.215
10	37, 38, 39	Automobiles, Aerospace, etc.	19	14.898
11	45, 48, 50, 53, 54	Miscellaneous	13	13.999
Total			184	

[a] Significant at the 5 percent level.

bias in favor of the neoclassical model may be present due to the Food and Beverage segment of the sample. With this possibility noted, we may consider the present results in light of those of JS.

Unlike the JS results, the enlarged comparisons reveal that only small differences can be identified in the overall relative importance of the liquidity, accelerator, and classical models, while a slight inferiority of the expected profits model to the other three is revealed. In the cross-sections, a clearer separation in the rankings appears, with the liquidity model being the most effective determinant of desired capital stock in nine of fifteen comparisons with the accelerator model, nine of thirteen comparisons with the expected profits model, and thirteen of fifteen comparisons with the neoclassical model. Similarly, the accelerator model is more effective than either the expected profits or neoclassical model while the neoclassical model is the least effective of those evaluated cross-sectionally.

TABLE 7—FOOD AND BEVERAGE FIRM COMPARISONS

	I_{cl}	I_l	I_a	I_{exp}
I_{cl}	—	14	17	15
I_l	12	—	15	15
I_a	9	11	—	15
I_{exp}	10	12	10	—

TABLE 8—COMPARISONS EXCLUDING FOOD AND BEVERAGES

	I_l	I_a	I_{cl}	I_{exp}
I_l	—	73	82	80
I_a	75	—	79	87
I_{cl}	67	71	—	82
I_{exp}	67	65	64	—

The relatively weak explanatory power of the expected profits proxy in our tests may shed some light on an earlier finding by Eisner (1967) that the rate of return on market value of the firm does not prove statistically significant in cross-sections and has a negative coefficient in firm-level time-series. Our $I_{\exp}$ variable is a measure of firm market value similar to the denominator in his expression. We find the total distributed lag effect of this variable to be inconsistent in direction in both time-series and cross-sections, and to be generally inferior in its explanatory significance. This suggests that Eisner's finding may be due to the lack of significance and consistency which attaches to the market-value measure in the denominator of his expression.

Furthermore, the relatively strong cross-sectional showing of the capacity-utilization-accelerator I_a variable relative to other desired capital stock proxies suggests a more Eisner-like and a less Jorgenson-like view of

corporate investment functions may be more appropriate than that indicated by the JS study. In fact, the accelerator proxy provided a better explanatory variable than the neoclassical model in nine of twelve cross-sectional comparisons. However, this view is not wholly reinforced by the time-series results which show the two to be virtually indistinguishable in the overall comparison and which show the accelerator explanation slightly superior to the neoclassical explanation (79 of 150 comparisions) in the homogeneity-corrected sample. Moreover, the rather high R^2 values for the cross-sectional tests of the I_a variable support Eisner's (1967) conclusion that the role of past sales changes in cross-sections is significant, presumably as a proxy for expected long-run pressures of demand on capacity.[8]

Based on strict numerical rankings, the results of the enlarged time-series and of the cross-sections agree in assigning the highest ranking to the liquidity model, I_l, and the second ranking to the accelerator model, I_a. For the third and fourth ranks, the neoclassical and expected profits models are reversed in the cross-sections from the time-series. With the earlier noted differences in the nature of cross-sectional and time-series evidence, the general correspondence in rankings suggests that the comparative effectiveness of the alternative models does not strongly depend upon the choice of time-series or cross-sectional perspective.

An additional comparison with JS can be obtained by following their procedure of enumerating the coefficients for the desired capital stock terms for each model. The frequency with which a desired capital stock term was included in an equation is taken by JS as an indication of its empirical importance. More significant desired capital stock proxies would lower the residual variation of the investment equation more frequently than less significant proxies, and would therefore be included and assigned nonzero values more frequently. Table 5 shows these results, along with those of JS. In the enlarged time-series and in the cross-sections, only small numerical differences separate the included coefficients for the four models, leaving an inconclusive picture on this criterion.

Taken together with the rankings of Table 4, the overall impression is that important differences in the empirical relevance of alternative models do not appear in our enlarged time-series sample. Some ability to distinguish among models derives from the cross-sections, and suggests that the liquidity explanation is the most effective alternative considered.

Thus, the major conclusion of the JS study that the neoclassical model of investment is more effective than other alternatives simply does not stand up to the enlarged test, and must now be interpreted as having no general implications beyond their sample. Other conclusions, such as their rejection of the liquidity model as a serious explanatory mechanism, must be similarly reinterpreted to apply to a much smaller set of firms. In the enlarged sample, the liquidity explanation of investment is among the most effective available.

The present finding is also of interest in light of the information content of the alternatives defined by JS. The accelerator, expected profits, and liquidity models are each functions of a single variable while the neoclassical model incorporates several variables including those comprising the accelerator and expected profits models. That the greater assemblage of information does not improve (and detracts from) the results obtained by the simpler models does not speak well for the empirical pertinence of the neoclassical formulation. Furthermore, the market value of the firm contained in both the expected profits and neoclassical formulations appears if anything to be an inferior means by which to explain investment behavior when employed either as an expectations or cost-of-capital measure.

Finally, evidence is accumulating which suggests that several of the variables involved in the JS comparison may be jointly determined along with investment as parts of a simultaneous system of corporate financial behavior. Empirical results by Mueller,

[8] Again, the differences in definition of sales change should be considered. See fn. 6.

Dhrymes and Kurz, and Elliott in particular lead in this direction. Under such circumstances, the single-equation structure of such a comparison may be inappropriate. Even putting this possible difficulty aside, the clearest indication obtained by the present results is that the most relevant explanatory model for individual corporate investment behavior is not the neoclassical framework of JS but an open question in need of more inquiry.

REFERENCES

P. J. Dhrymes, and M. Kurz, "Investment, Dividend, and External Finance Behavior of Firms," in *Determinants of Investment Behavior,* Universities—Nat. Bur. Econ. Res. Conference series, New York 1967.

R. Eisner, "Tax Policy and Investment Behavior: Further Comment," *Amer. Econ. Rev.,* Sept. 1970, *60,* 746–51.

———, "Investment and the Frustrations of Econometricians," *Amer. Econ. Rev. Proc.,* May 1969, *59,* 50–64.

———, "A Permanent Income Theory for Investment: Some Empirical Explorations," *Amer. Econ. Rev.,* June 1967, *57,* 363–90.

——— **and M. I. Nadiri,** "Investment Behavior and the Neoclassical Theory," *Rev. Econ. Statist.,* Aug. 1968, *50,* 369–82.

J. W. Elliott, "Forecasting and Analysis of Corporate Financial Performance with an Econometric Model of the Firm," *J. Finance Quant. Anal.,* Mar. 1972, *7,* 1499–1526.

Y. Grunfeld, "The Determinants of Corporate Investment," in A. C. Harberger, ed., *The Demand for Durable Goods,* Chicago, London 1960, 211–66.

D. W. Jorgenson, "Capital Theory and Investment Behavior," *Amer. Econ. Rev. Proc.,* May 1963, *53,* 247–59.

——— **and R. E. Hall** "Tax Policy and Investment Behavior," *Amer. Econ. Rev.,* June 1967, *57,* 391–414.

——— **and C. D. Siebert,** (1968a) "Theories of Corporate Investment Behavior," *Amer. Econ. Rev.,* Sept. 1968, *58,* 681–712.

——— **and** ——— (1968b) "Optimal Capital Accumulation and Corporate Investment Behavior," *J. Polit. Econ.,* Nov./Dec. 1968, *76,* 1123–51.

E. Kuh, *Capital Stock Growth: A Microeconometric Approach,* Amsterdam 1954.

D. Mueller, "The Firm Decision Process: An Econometric Investigation," *Quart. J. Econ.,* Feb. 1967, *81,* 58–87.

Standard and Poor's Compustat Tape 1949–1968, Standard Statistics Division, New York.

Unemployment and Consumption: Note

By MICHAEL GROSSMAN*

According to a new approach to demand theory, households produce the commodities that enter their utility functions with inputs of market goods and their own time (see Gary Becker, Becker and Robert T. Michael, Kelvin Lancaster, Richard Muth). Provided time and goods were not employed in fixed proportions in the production of commodities, consumers would have an incentive to substitute the former for the latter if the price of their time, measured, say, by their potential real wage rate, declined. Since the incidence of unemployment causes the price of time to fall, one would expect unemployed workers to substitute their own time for market goods in the production of commodities.

In this note, I use the notion of substitution in production between goods and time to interpret data on detailed family consumption expenditures before and during unemployment. The information is contained in a survey of the insured unemployed. In particular, I focus on differences in the pattern of expenditure reductions that emerge when primary market workers are unemployed, compared to when secondary market workers are unemployed. I also examine differences in the estimated income elasticity of total consumption for claimants in low unemployment areas compared to that for claimants in high unemployment areas. A tentative part of this analysis is a computation of the reduction in total consumption due to substitution of time for goods that would occur if unemployment were fully anticipated and, hence, not accompanied by a reduction in permanent income.

I. Data and Methods

This note is based on data from six studies of the insured unemployed, conducted by the Bureau of Employment Security (*BES*) of the U.S. Department of Labor between 1953 and 1958.[1] The labor market areas in the *BES* sample, the dates during which each survey was conducted, and the overall unemployment rate of the area during those dates are shown in Table 1. The surveys included average monthly family expenditures before and during unemployment for eleven exhaustive components of total consumption. Data were collected from three types of claimants: single individuals, heads of four person households (primary market workers), and nonheads of four person households (secondary market workers).[2] There were 462 single individuals in the sample, 774 heads of four person households, and 482 nonheads. In the *BES* publication based on the surveys, the expenditure information was given in the form of area averages by claimant type. Hence, there were six observations for each claimant type.

I deflated outlays in each expenditure category by a corresponding component of the Consumer Price Index (*CPI*) (1957—59 = 100). I selected the appropriate annual deflators for before- and during-unemployment expenditures as follows. For Pittsburgh, the unemployment period occurred mainly during 1954. Therefore, deflation of outlays during this period made use of the 1954 *CPI* components. The deflators used with before-unemployment spending were chosen by applying the average duration of unemploy-

* Research associate, National Bureau of Economic Research. I am grateful to Gary S. Becker, Carol Breckner, Ilene J. Grossman, Arlene Liebowitz, Robert T. Michael, Margaret Reid, Paul Wachtel, an anonymous referee, and the editor of this *Review* for their helpful comments on an earlier draft.

[1] For a complete description of the data, see *BES* No. U-203. Note that the *BES* is now part of the Manpower Administration.

[2] Although the published data did not give the sex of the claimant, most nonheads of four person households are undoubtedly wives. Since differences in the economic roles of primary and secondary market workers in the household can explain my main results, I do not emphasize differences based on a sex classification in the paper. I would like to thank Carol Breckner, Ilene J. Grossman, and a referee for urging me to adopt the primary worker-secondary worker terminology. This terminology should prevent readers from getting the mistaken impression that I am a "male chauvinist pig."

TABLE 1—THE *BES* SURVEY

Area	Dates	Unemployment Rate (percent)
Pittsburgh, Pennsylvania	9/53–8/54	9.4
Tampa-St. Petersburg, Florida	11/55–10/56	4.2
Andersonville-Greenville-Spartanburg, South Carolina	4/56–3/57	3.7
Albany-Schenectady-Troy, New York	5/56–4/57	3.4
Portland, Oregon	4/57–3/58	10.0
St. Louis, Missouri	4/57–3/58	9.2

ment to the midpoint date of the survey period. For Pittsburgh, the midpoint date was February 15, 1954. Since claimants reported an average duration of unemployment of 20 months, it was assumed to begin in June 1952, and outlays while employed were deflated by the *CPI* components for 1952. A similar procedure was applied to the expenditure data of claimants in the other five areas.

After deflation, I computed mean expenditures before and during unemployment by claimant type. The mean expenditure on a given item was defined as a weighted average of the six area means, where the weights were the proportion of the sample in each area. Note that my procedure eliminated the effects on the expenditure data of price changes both during the entire 1953–58 period and during the movement from employment to unemployment in each area.

II. Differences in Head-Nonhead Expenditure Reductions

Table 2 gives consumption elasticities and marginal propensities to consume for four person families in which the head was unemployed, and for four person families in which the nonhead was unemployed. The consumption elasticity of food, for example, equals the percentage reduction in deflated family food outlays caused by unemployment divided by the percentage reduction in deflated total consumption. The marginal propensity to consume food equals the absolute reduction in deflated food outlays divided by the absolute reduction in deflated total consumption.

In comparing families in which the head was unemployed to those in which the nonhead was unemployed, I use the reduction in current consumption as an index of the reduction in permanent income due to unemployment. The reduction in current consumption is used instead of the reduction in current income because the former clearly has a smaller transitory component than the latter. Since to a large extent, unemployment causes a temporary decline in income, one

TABLE 2—CONSUMPTION ELASTICITIES AND MARGINAL PROPENSITIES TO CONSUME FOR FOUR PERSON HOUSEHOLDS WHERE HEAD OR NONHEAD IS UNEMPLOYED

	Consumption Elasticity		Marginal Propensity to Consume	
Item	Head	Nonhead	Head	Nonhead
Food	.74	.74	.22	.20
Housing and utilities	.79	.35	.17	.07
Household operations	.97	4.16	.05	.27
House furnishings	1.03	.62	.04	.02
Clothing	1.61	2.50	.09	.15
Medical care	1.45	.54	.07	.02
Transportation	1.14	.88	.17	.15
Personal care	.00	1.39	.00	.02
Tobacco and alcoholic beverages	.97	.89	.04	.02
Reading and recreation	.91	1.31	.04	.05
Other	1.77	.32	.11	.03

Note: Computed from: U.S. Department of Labor, *BES*, Tables B-2, B-6, D-3, and D-5.

would expect persons to react to it mainly by adjusting their savings.[3]

According to traditional demand theory, the various goods and services in Table 2 are the basic arguments in consumers' utility functions. Provided relative prices did not change when workers became unemployed, it would attribute reductions in the consumption of these goods and services solely to an income effect. If families in which the head was unemployed had the same set of permanent income or consumption elasticities as families in which the nonhead was unemployed, then the correlation coefficient between these two sets of elasticities should equal +1. In fact, the Spearman rank correlation coefficient between the consumption elasticities in Table 2 is −.23, and the product-moment correlation coefficient is −.04.[4] Such low (and even negative) correlations are inconsistent, at least at first glance, with the hypothesis that the two claimant groups have the same set of permanent income elasticities.

Some of the differences in the pattern of expenditure reductions revealed by Table 2 are striking. Whether the dissimilarity between the two consumption elasticities for each item in the table is measured by their ratio or by their difference, household operations and clothing deserve special mention. Their consumption elasticities for families in which the nonhead was unemployed are, respectively, four times as large and one and a half times as large as the corresponding elasticities for families in which the head was unemployed. In terms of marginal propensities to consume, the differential impact of unemployment of heads and nonheads on the quantity of household operations demanded is most noticeable. The reduction in household operations for nonheads accounts for 27 percent of the reduction in total consumption. The corresponding figure for heads is 5 percent.[5]

The household production function approach to demand analysis provides a cogent explanation of the relatively large reductions in the quantities of household operations and clothing demanded when secondary market workers become unemployed. Its explanation is based on the proposition that the elasticities of derived demand for various market goods with respect to the price of secondary workers' time are likely to exceed the elasticities with respect to the price of primary workers' time.

For instance, the specific items included in household operations outlays are inputs, together with primary workers' time and secondary workers' time, into the production of commodities such as "cleanliness of the home," "cleanliness of the people who live in the home," and "child care."[6] When either primary workers or secondary workers become unemployed, the price of their time falls, and an incentive is created to substitute this time for market goods in the production of cleanliness or child care. Since secondary market workers are primary nonmarket workers, the share of their time in the total cost of producing certain commodities is likely to exceed the share of primary workers' time in total cost. Thus, if the Allen partial elasticities of substitution in production (see R. G. D. Allen, pp. 503–09) between primary workers' time and market goods and between secondary market workers' time and market goods were equal, the derived demand elasticity of household operations with respect to the price of secondary workers' time would be greater.[7] To the extent that the secondary

[3] If the permanent income elasticity of total consumption equaled unity, then the percentage change in consumption would coincide with the percentage change in permanent income.

[4] If the category Other Consumption is omitted from these calculations, the rank correlation coefficient is +.02, and the product-moment correlation coefficient is +.12.

[5] Although the consumption elasticity of personal care is 1.4 for nonheads and 0 for heads, the differential impact of unemployment of heads and nonheads on the quantity of this item demanded falls dramatically when marginal propensities to consume are examined.

[6] Household operations outlays consist of 1) laundry sent out, 2) cleaning of rugs and draperies, 3) launderette and coin operated washing machine service, 4) wages and tips to domestic servants, 5) laundry soaps and cleaning supplies, and 6) fees to day nurseries, child care centers, and baby sitters.

[7] In analyzing derived demand elasticities, I rule out substitution in consumption among commodities by as-

worker partial elasticity of substitution exceeded the primary worker partial elasticity of substitution, the difference in derived demand elasticities would be even larger.

Along similar lines, clothing and time produce commodities such as "warmth" and "style." Once again, the results in Table 2 are consistent with the hypothesis of a positive differential between the shares of secondary and primary workers' time in total cost. Presumably, in some household activities, there are negative differentials between these shares or greater partial elasticities of substitution between primary workers' time and market goods. Indeed, for the items housing and utilities, other consumption, and medical care, head consumption elasticities and marginal propensities to consume exceed the corresponding nonhead estimates. The first two items, however, consist of diverse components, and the medical care differential can be explained by the purchase of group health insurance by the head for himself and his family at his place of work. Therefore, it is not possible to use the *BES* budget studies to identify household activities in which primary workers' time and market goods are good substitutes.

It should be noted that within the context of the household production function model, the low correlation between the consumption elasticities of heads and nonheads need not imply that the two groups have different "pure" permanent income elasticities. By pure permanent income elasticity, I mean the income elasticity that would be observed if a change in permanent income were not accompanied by a change in the value of the time of one or more family members. This would occur if, for example, a family, all of whose members worked in the market, suffered an unanticipated decline in property income. Since unemployment lowers the value of time, it generates substitution as well as income effects. These substitution effects can go a long way in explaining the differential impact of the unemployment of primary market workers and secondary market workers on the quantities of various market goods demanded.

III. Income Elasticities of Total Consumption

Three of the *BES* studies were undertaken during years of business contraction, while the other three were conducted during years of relative prosperity (see Table 1). The three labor market areas that were surveyed during recessions had much higher unemployment rates than the other three. In this section, I compare the income elasticity of total consumption for claimants in high unemployment areas to that for claimants in low unemployment areas. Presumably, claimants in high unemployment areas were unemployed mainly for cyclical reasons. On the other hand, unemployment in the other areas was presumably due to seasonal or frictional factors.[8]

In general, it is much more difficult for workers to predict or anticipate cyclical unemployment than to predict seasonal or frictional unemployment. Hence, they would view a loss in income due to the former type of unemployment as a more serious loss than that due to the latter type. Based on this argument, unemployed workers in high unemployment areas should exhibit larger income elasticities of total consumption than those in low unemployment areas. Table 3 illustrates the truth of this proposition. For each of the three claimant types, the estimated income elasticity of consumption in high unemployment areas is approximately twice as large as the income elasticity in low unemployment areas.[9]

suming that *relative* prices of commodities are not affected by changes in the price of time. In other words, I assume that the shares of secondary and primary workers' time in the total cost of producing a given commodity do not vary among commodities. Given the fairly aggregate nature of the items in Table 2, this is a weak assumption.

[8] Since all of the members of the sample were receiving unemployment compensation benefits, it is not likely that they were unemployed for long-term or structural reasons.

[9] The income elasticity of total consumption equals the percentage reduction in deflated family consumption due to unemployment divided by the percentage reduction in deflated family income. Note that family income during unemployment includes unemployment compensation benefits.

TABLE 3—INCOME ELASTICITIES OF TOTAL CONSUMPTION BY CLAIMANT TYPE AND UNEMPLOYMENT LEVEL OF AREA

Group	Income Elasticity
Single individuals	
High unemployment areas	.38
Low unemployment areas	.23
Heads of four person households	
High unemployment areas	.46
Low unemployment areas	.20
Nonheads of four person households	
High unemployment areas	.43
Low unemployment areas	.18

If seasonal or frictional unemployment were fully anticipated, it would cause no reduction in permanent income. This does not mean that such unemployment would have no impact on consumption. The price of time of seasonally unemployed workers would decline temporarily, and they would have an incentive to substitute time for goods in the production of commodities.[10]

By assuming that workers in low unemployment areas fully anticipated their unemployment, it is possible to put an upper limit on the reduction in consumption due to substitution of time for goods. If the income elasticity for heads is employed in this calculation, then a 10 percent reduction in family income due to fully anticipated unemployment would reduce consumption by 2 percent on account of the substitution effect. Put differently, the income elasticity for heads in high unemployment areas shows that if family income fell by 10 percent in a recession, consumption would fall by 4.6 percent. A little less than half of this decline would be due to the substitution of time for goods, and a little more than half would be due to the reduction in permanent income. It should be emphasized that this is an extremely tentative calculation. To the extent that workers in low unemployment areas did not anticipate their unemployment, part of their reduction in consumption would be due to an income effect. I believe, however, that the incentive to substitute time for goods during periods of unemployment can explain at least part of the decline in total consumption during such periods.

[10] This point has been made by Becker and Gilbert R. Ghez. Note from Table 3 that with the level of unemployment held constant, income elasticities are practically the same for each of the three types of claimants. Therefore, there is no evidence in the *BES* data that secondary market workers' time and market goods are better substitutes in *all* household activities than primary market workers' time and market goods.

IV. Summary and Implications

In this note, I have argued that in the production of the household commodities, cleanliness child care, warmth, and style, secondary market workers' time and market goods are better substitutes than primary market workers' time and market goods. This explains the relatively large declines in outlays on household operations and clothing that are observed when secondary workers become unemployed. During recent decades, the labor force participation rates of married women and other secondary market workers have increased substantially. Therefore, an immediate implication of my findings is that the pattern of cyclical fluctuations in the various components of total consumption might change in the future. Given the relatively high elasticities of derived demand for various market goods with respect to changes in the price of secondary market workers' time, cyclical fluctuations in household operations and in clothing should increase in the future.

In this note, I have also argued that not all of the reduction in total consumption that accompanies unemployment can be attributed to an income effect. Instead, part of this reduction is due to the incentive to substitute time for goods in the production of commodities. Hence, even if unemployment compensation benefits offset a substantial fraction of any decline in permanent income caused by unemployment, one would expect total consumption to fall during such periods.

REFERENCES

R. G. D. Allen, *Mathematical Analysis for Economists*, London 1962.

G S. Becker, "A Theory of the Allocation of Time," *Econ. J.*, Sept. 1965, *75*, 493–517.

——— and G. R. Ghez, "The Allocation of

Time and Goods Over the Lifetime," mimeo., Nat. Bur. Econ., Res., New York 1971.

——— and R. T. Michael, "On the Theory of Consumer Demand," unpublished 1970.

K. J. Lancaster, "A New Approach to Consumer Theory," *J. Polit. Econ.*, Apr. 1966, *74*, 132–57.

R. Muth, "Household Production and Consumer Demand Functions," *Econometrica*, July 1966, *34*, 699–708.

U.S. Department of Labor, Bureau of Employment Security, *Unemployment Insurance and the Family Finances of the Unemployed: An Analysis of Six Benefit Adequacy Studies, 1953–58*, *BES* No. U-203, Washington 1962.

A Test for Relative Economic Efficiency: Some Further Results

By Pan A. Yotopoulos and Lawrence J. Lau*

Economic efficiency can be decomposed into two components: technical efficiency and price efficiency. A firm is said to be more technically efficient than another if it consistently produces larger quantities of output from the same quantities of measurable inputs. Differences in technical efficiency among (groups of) firms can therefore be captured by quantifying their differences in "technology."[1] On the other hand, a firm is said to be price-efficient if it maximizes profits. Since profit maximization implies equalization of the value of the marginal product of each variable input to its price, analysis of price efficiency involves also variables that are not ordinarily included in production function analysis—prices.[2] Differences in economic efficiency among (groups of) firms may be caused by differences in technical and/or price efficiency.

We compare relative economic efficiency of two firms with varying degrees of technical and price efficiency. For this purpose we compare the actual profit functions of the two firms, at given output and input prices and quantities of measurable fixed inputs. The firm which has higher profits (i.e., total revenue minus the total cost of the variable factors of production) within a certain specified range of output and input prices is considered to be the more economic-efficient firm (within that range of prices). By applying this concept in the comparison of small and large Indian farms, we found in our previous paper (see Lau and Yotopoulos (1971)) that small farms are more economic-efficient. For many purposes, however, it is important to assess the relative importance of the two factors responsible for observable differences in relative economic efficiency. In this paper, we extend our earlier work with the specific purpose of identifying and isolating the causes for such difference between small and large farms. As an illustration we apply our technique to the same body of data as used in Lau and Yotopoulos (1971).

Knowledge of the values of the technical- and price-efficiency parameters may be crucial in the formulation of policies. Suppose that differences in relative economic efficiency stem from differences in behavioral decision rules associated with the ability to maximize profits, then identifying the two components of technical and price efficiency can lead to better predictions. Again, suppose it were socially desirable to bring about the "value of the marginal product equals cost" rule, it would be possible, if the actual decision rules are known, to work out a system of ad valorem taxes and subsidies on the variable inputs and lump sum taxes and subsidies so that the social optimum would be attained. Finally, such knowledge may have important implications on agrarian reform, agricultural education, and agricultural extension policies.

As a by-product of our analysis we also perform a simple statistical test of the hypothesis of constant returns to scale within the framework of the profit function. In our specific application the question of whether agriculture is characterized by increasing returns, constant, or decreasing returns is a

* Associate professor, Food Research Institute, and assistant professor, department of economics, Stanford University, respectively. Earlier drafts of this paper benefitted from comments of Richard R. Nelson and Vernon W. Ruttan. Wuu-Long Lin provided competent research assistance. The research was partially supported by National Science Foundation grant GS-2874-A1 to Stanford University. We alone are responsible for any errors therein.

[1] Examples of studies that focus on technical efficiency are Michael Farrell, Yair Mundlak, W. D. Seitz (1970), Robert Solow, and C. Peter Timmer (1970).

[2] Examples of studies of price efficiency are W. David Hopper, Theodore W. Schultz, and Yotopoulos (1967a, 1968). These studies, however, are more limited than our approach to price efficiency since they allow only for inter-group (instead of both inter- and intra-group) variations in factor prices. The difference is further spelled out in J. Wise and Yotopoulos (1969).

crucial point that has important policy implications.

In Section I, the model is briefly described. In Section II, the relevant statistical tests are formulated in the context of the Cobb-Douglas function. Section III is on empirical implementation and statistical results. Section IV contains a brief summary and conclusions.

I. The Model

The model that we use is identical to the one in Lau and Yotopoulos (1972). The basic extension that allows us to identify separately the components of technical and price efficiency lies in the utilization of the factor demand functions. Let there be two firms[3] denoted by the superscripts, with the production functions, respectively,[4]

$$(1)\quad V^1 = A^1F(X^1, Z^1);\; V^2 = A^2F(X^2, Z^2)$$

where V is output, X and Z are vectors representing the variable inputs and the fixed inputs of production, respectively. Equation (1) incorporates firm-specific technical efficiency,[5] captured by the neutral differences in the production function, A^1 and A^2. Two firms are equally technical-efficient if, and only if, $A^1=A^2$.

The marginal conditions are given by

$$(2)\quad \begin{aligned} &\frac{\partial A^1F(X^1, Z^1)}{\partial X_j^1} = k_j^1 c_j^1 \\ &k_j^1 \geqq 0 \\ &\frac{\partial A^2F(X^2, Z^2)}{\partial X_j^2} = k_j^2 c_j^2 \\ &k_j^2 \geqq 0; \\ &j = 1, \ldots, m \end{aligned}$$

where c_j^i is the money price of the jth input divided by the money price of output faced by the ith firm. The c_j^i's will be referred to as normalized prices. Equation (2) allows a firm to be unsuccessful in its attempts to equate values of the marginal product of its inputs to their respective normalized prices. This is introduced through the firm-specific and variable input-specific k_j^i and it represents differences in managerial-entrepreneurial ability. If, and only if, two firms are equally price-efficient with respect to all variable inputs, then $k_j^1=k_j^2$, for $j=1, \ldots, m$.[6]

To the production function $F(X, Z)$ corresponds a Unit-Output-Price (UOP) profit function,[7]

$$(3)\quad \Pi = G(c_1, \ldots, c_m; Z_1, \ldots, Z_n)$$

where Π is UOP profit (i.e., current revenues less current total variable costs) divided by the money price of output and c_j is the normalized price of variable input X_j. As shown in our 1971 article, the actual demand functions are given by

$$(4)\quad X_j^i = \frac{-A^i}{k_j^i}\,\frac{\partial G(k^ic^i/A^i; Z^i)}{\partial c_j^i},\qquad i = 1, 2; j = 1, \ldots, m$$

The supply function is given by

$$(5)\quad V^i = A^i\left(G(k^ic^i/A^i; Z^i) - \sum_{j=1}^{m} c_j^i\,\frac{\partial G(k^ic^i/A^i; Z^i)}{\partial c_j^i}\right)\qquad i = 1, 2; j = 1, \ldots, m$$

[3] This analysis can be easily extended to n firms.

[4] In the notation that follows, we will continue at times to use the variables X, Z, k, and c unsubscripted in order to denote vectors.

[5] Consistent with the "technical progress" literature, these parameters represent differences in environmental factors, in engineering entrepreneurial ability and in other nonmeasurable fixed factors of production.

[6] In our formulation, k_j reflect a general systematic rule of behavior—a decision rule that gives the profit-maximizing marginal productivity conditions as a special case. That the decision rule for the firm consists of equating the marginal product to a constant times the normalized price of each input may be rationalized as follows: (i) Consistent over- or under-valuation of the opportunity costs of the resources by the firm; (ii) Satisficing behavior; (iii) Divergence of expected and actual normalized prices; (iv) Divergence of the subjective probability distribution of the normalized prices from the objective distribution of normalized prices; (v) The elements of k^i may be interpreted as the first-order coefficients of a Taylor's series expansion of arbitrary decision rules of the type (over)

From the actual supply and demand functions, we can obtain the *actual UOP profit function*,

$$(6)\quad \Pi_a^i = V^i - \sum_{j=1}^{m} c_j^i X_j^i = A^i\left(G(k^i c^i/A^i; Z^i) + \sum_{i=1}^{m} \cdot \frac{(1-k_j^i)}{k_j^i}\frac{\partial G(k^i c^i/A^i; Z^i)}{\partial c_j^i}\right), \quad i = 1, 2; j = 1, \ldots, m$$

Note that even if there exists no difference in technical efficiency (i.e., the A^i), there may be more than one set of k^i's that correspond to the same actual UOP profit function.

It should be emphasized at this point that X_j^i, V^i, and Π_a^i as given in (4), (5), and (6) are the actual quantities of inputs demanded, output supplied, and UOP profits received by firm i given the firm-specific A^i and the firm- and input-specific k_j^i. When appropriate functional forms are specified for G in equation (3), then equations (4)–(6) may be derived and estimated. Note that because of the profit identity only $(m+1)$ of the $(m+2)$ equations in (4), (5), and (6) are independent. Hence only $(m+1)$ of the functions need be estimated.

Observe that: (i) $\partial\Pi_a^i/\partial A^i > 0$, i.e., actual profit always increases with the level of technical efficiency for given normalized input prices and k^i; (ii) When $k_j^i = 1$ for $j=1, \ldots, m$, the firm is a true profit maximizer; (iii) If, and only if, $A^1 = A^2$ *and* $k^1 = k^2$ the actual UOP profit functions and the demand functions of the two firms coincide with each other. From these observations a number of tests for relative economic efficiency become possible.

First, one can test the null hypothesis of equal relative economic efficiency. This hypothesis is equivalent to testing whether there exists significant differences between the profit functions. Such a test has been carried out in our 1971 article, in which the hypothesis of equal relative economic efficiency is rejected.

Second, it is possible to test separately the hypotheses of equal technical efficiency, i.e., $A^1 = A^2$, and of equal price efficiency, i.e., $k^1 = k^2$, or both. Note that one can have equal relative economic efficiency without necessarily having both $A^1 = A^2$ and $k^1 = k^2$. Finally, one can compute A^i and k^i explicitly to isolate the sources of differences in economic efficiency. These are the specific tasks of this paper.

II. The Formulation of the Cobb-Douglas Case

A Cobb-Douglas production function with decreasing returns in the m variable inputs and with n fixed inputs is given by[8]

$$V = A \prod_{j=1}^{m} X_j^{\alpha_j} \prod_{j=1}^{n} Z_j^{\beta_j}$$

where

$$\mu \equiv \sum_{j=1}^{m} \alpha_j < 1$$

The UOP profit function for this Cobb-Douglas production function is

$$(7)\quad \Pi^* = A^{(1-\mu)^{-1}}(1-\mu) \cdot \left(\prod_{j=1}^{m}\left(\frac{c_j}{\alpha_j}\right)^{-\alpha_j(1-\mu)^{-1}}\right) \cdot \left(\sum_{j=1}^{n} Z_j^{\beta_j(1-\mu)^{-1}}\right)$$

By direct computation, the actual UOP profit function for this Cobb-Douglas production function for firm i, with efficiency parameters A^i and k^i is

$$\Pi_a^i = A^{i(1-\mu)^{-1}}\left(1 - \sum_{j=1}^{m} \alpha_j/k_j^i\right) \cdot \left(\prod_{j=1}^{m} (k_j^i)^{-\alpha_j(1-\mu)^{-1}}\right) \cdot \left(\prod_{j=1}^{m} \alpha_j^{\alpha_j(1-\mu)^{-1}}\right)$$

$$\frac{\partial F}{\partial X_j^i} = f_j^i(c_j^i), \qquad i = 1, 2,$$

where $f_j^i(0) = 0$ and $f_j^{i\prime}(c_j^i) \geq 0$. A wide class of decision rules may be encompassed under this formulation.

[7] For a more detailed discussion of these derivations, see our 1971 article.

[8] The value of $\mu < 1$ is required since constant or increasing returns in the variable inputs are inconsistent with profit maximization.

$$\cdot\left(\prod_{j=1}^{m}(\overset{i}{c_j})^{-\alpha_j(1-\mu)^{-1}}\right)$$

$$\cdot\left(\prod_{j=1}^{n}(\overset{i}{Z_j})^{\beta_j(1-\mu)^{-1}}\right) \qquad i=1,2$$

or

$$\text{(8)}\quad \overset{i}{\Pi_a} = \overset{i}{A_*}\left[\prod_{j=1}^{m}(\overset{i}{c_j})^{\alpha_j^*}\right]\left[\prod_{j=1}^{n}(\overset{i}{Z_j})^{\beta_j^*}\right] \qquad i=1,2$$

where

$$\overset{i}{A_*} \equiv \overset{i}{A}^{(1-\mu)^{-1}}\left(1-\sum_{j=1}^{m}\alpha_i/\overset{i}{k_j}\right)$$

$$\cdot\left(\prod_{j=1}^{m}(\overset{i}{k_j})^{-\alpha_j(1-\mu)^{-1}}\right)$$

$$\cdot\left(\prod_{j=1}^{m}\alpha_j^{\alpha_j(1-\mu)^{-1}}\right) \qquad i=1,2$$

$$\text{(9)}\quad \overset{i}{k_*} \equiv \left(1-\sum_{j=1}^{m}\alpha_j/\overset{i}{k_j}\right)(1-\mu)^{-1} \qquad i=1,2$$

$$\overset{*}{\alpha_j} \equiv -\alpha_j(1-\mu)^{-1} < 0 \qquad j=1,\ldots,m$$

$$\overset{*}{\beta_j} \equiv \beta_j(1-\mu)^{-1} > 0 \qquad j=1,\ldots,n$$

Similarly by direct computation the demand functions for variable inputs are given by

$$\overset{i}{X_l} = \overset{i}{A}^{(1-\mu)-1}(\alpha_l/\overset{i}{k_l}\overset{i}{c_l})\left(\sum_{j=1}^{m}(\overset{i}{k_j})^{-\alpha_j(1-\mu)^{-1}}\right)$$

$$\cdot\left(\prod_{j=1}^{m}(\alpha_j)^{\alpha_j(1-\mu)^{-1}}\right)\left(\prod_{j=1}^{m}(\overset{i}{c_j})^{-\alpha_j(1-\mu)^{-1}}\right)$$

$$\cdot\left(\prod_{j=1}^{n}(\overset{i}{Z_j})^{\beta_j(1-\mu)^{-1}}\right) \qquad i=1,2;\quad l=1,\ldots,m$$

$$\text{(10)}\quad \overset{i}{X_l} = -\overset{i}{A_*}\overset{*}{\alpha_l}(\overset{i}{k_l})^{-1}(\overset{i}{c_l})^{-1}(\overset{i}{k_*})^{-1}$$

$$\cdot\left(\prod_{j=1}^{m}(\overset{i}{c_j})^{\alpha_j^*}\right)\left(\prod_{j=1}^{n}(\overset{i}{Z_j})^{\beta_j^*}\right) \qquad i=1,2;\quad l=1,\ldots,m$$

or, by substitution from equation (8)

$$\text{(11)}\quad \frac{-\overset{i}{c_l}\overset{i}{X_l}}{\overset{i}{\Pi_a}} = (\overset{i}{k_l})^{-1}(\overset{i}{k_*})^{-1}\overset{*}{\alpha_l} \equiv \overset{*i}{\alpha_l} \qquad i=1,2;\quad l=1,\ldots,m$$

Equation (8) indicates that the actual *UOP* profit functions of the two firms differ by a constant factor, which is a function of k_j^i and A^i. In addition, from equation (11) all the demand functions differ across firms by constant factors. The basic estimating equations for this study are (8) after taking natural logarithms, and (11).

The first hypothesis we can investigate within our framework is that of equal relative economic efficiency. This implies that

$$\text{(12)}\quad \overset{1}{A_*} = \overset{2}{A_*}$$

or that $ln\ (A_*^1/A_*^2)=0$. Note that it is possible for two firms to be equally economic-efficient without being equally technical-efficient or equally price-efficient or both. We note, moreover, that if, and only if, $A^1=A^2$ *and* $k^1=k^2$, then $A_*^1=A_*^2$ *and* $\alpha_l^{*1}=\alpha_l^{*2}$, $l=1,\ldots,m$ and the two functions Π_a^1 and Π_a^2 (or $ln\ \Pi_a^1$ and $ln\ \Pi_a^2$) should be identical. This implies the following linear relations:

$$\text{(13)}\quad \text{(i)}\ \overset{1}{A_*} = \overset{2}{A_*} \quad \textit{and}$$

$$\text{(ii)}\ \overset{*1}{\alpha_l} = \overset{*2}{\alpha_l}, \qquad l=1,\ldots,m$$

We can therefore test for the two components of economic efficiency, i.e., for equal *relative technical* and *price efficiency* by testing statistically the joint hypotheses in (13). Furthermore, we observe that if, and only if, $k^1=k^2$, then

$$\text{(14)}\quad \overset{*1}{\alpha_l} = \overset{*2}{\alpha_l}, \qquad l=1,\ldots,m$$

Hence a test of equal *relative price efficiency* (but not equal economic efficiency) consists of testing the joint hypotheses in (14).

In our model we may also investigate the *absolute price efficiency* hypothesis, i.e., whether one or both firms have maximized

profits subject to given prices. Observe that perfect profit maximization in the ith firm or group of firms implies

$$(15) \quad \text{(i) } k_l^i = 1 \qquad l = 1, \ldots, m \quad \text{and}$$

$$\text{(ii) } k_*^i = 1$$

From (15) it follows that

$$(16) \qquad \overset{*i}{\alpha_l} = \overset{*}{\alpha_l}, \qquad l = 1, \ldots, m$$

This furnishes a necessary and sufficient statistical test of the hypothesis of profit maximization.[9]

Next we turn our attention to the question of the degree of returns to scale. For the case of a Cobb-Douglas profit function, the necessary and sufficient condition for homogenity of degree k of the underlying production function is [10]

$$\frac{(k-1)}{k}\sum_{j=1}^{m}\overset{*}{\alpha_j} + \frac{1}{k}\sum_{j=1}^{n}\overset{*}{\beta_j} = 1$$

or alternatively,

$$\sum_{j=1}^{n}\overset{*}{\beta_j} = k - (k-1)\sum_{j=1}^{m}\overset{*}{\alpha_j}$$

[9] To see that (16) implies (15), and hence is a necessary and sufficient condition for profit maximization for firm i, just observe that (16) implies

$$k_l^{i-1} k_*^{i-1} = 1, \qquad l = 1, \ldots, m$$

which in turn implies that

$$k_l^{i-1} = k_{l'}^{i-1}, \qquad l, l' = 1, \ldots, m$$

Hence

$$k_l^{i-1} = k_*^i = \frac{\left(1 - k_l^{i-1}\sum_{j=1}^{m}\alpha_j\right)}{\left(1 - \sum_{j=1}^{m}\alpha_j\right)}$$

which leads to

$$k_l^{i-1} = 1 \qquad l = 1, \ldots, m$$

Equation (16) therefore furnishes a strong statistical test of the hypothesis of profit maximization.

[10] This result is derived in Lau (1972) and Lau and Yotopoulos (1972).

Note that $\sum_{j=1}^{m}\alpha_j^* < 0$ by the monotonicity conditions on the profit function. Hence, if $k > 1$ (increasing returns), $\sum_{j=1}^{n}\beta_j^* > 1$. If $k = 1$ (constant returns), $\sum_{j=1}^{n}\beta_j^* = 1$. If $k < 1$ (decreasing returns), $\sum_{j=1}^{n}\beta_j^* < 1$. A test of the hypothesis of constant returns *in all inputs* in the Cobb-Douglas case then becomes a test of the hypothesis $\sum_{j=1}^{n}\beta_j^* = 1$, where β_j^* are the elasticities of the profit function with respect to the fixed factors of production. We shall test this hypothesis in our application.

Finally, based on the definitional identities in equations (9) and (11) and the estimates of $ln\ A_*^i$, α_j^*, β_j^*, and α_l^{*i}, one can compute the values of A^i and the k^i, as well as the direct production elasticities of the inputs α_j and β_j.

III. Empirical Analysis

In this section we use the same data as used in our 1971 article from the *Studies in the Economics of Farm Management* (*The Studies*) of the Indian Ministry of Food and Agriculture to estimate jointly the *UOP* profit function and the labor demand function conditional on the given quantities of cultivable land and fixed capital.[11]

The profit function (8) and the labor demand function (11) may be written

$$(17) \quad ln\,\Pi_a = ln\,\overset{S}{A_*} + \overset{*}{\delta_L} D_L + \overset{*}{\alpha_1}\, ln\, w + \overset{*}{\beta_1}\, ln\, K + \overset{*}{\beta_2}\, ln\, T$$

$$(18) \quad \frac{-wL}{\Pi_a} = \overset{*L}{\alpha_1} D_L + \overset{*S}{\alpha_1} D_S$$

where Π_a is actual *UOP* profit (total revenue less total variable cost, divided by the price of output), w is normalized wage rate, K is interest on fixed capital, T is cultivable land, $\delta_L \equiv ln\,(A_*^L/A_*^S)$, and D_L and D_S are dummy variables taking the value of one for large and small farms, respectively, and zero otherwise. Large farms are defined as those with cultivable land greater than ten acres per farm. A maintained hypothesis is that the production function is identical on large and

[11] A description of the data is available in our 1971 article. We shall not repeat it here.

small farms up to a neutral efficiency parameter. This implies that the coefficients corresponding to $ln\ w$, $ln\ K$, and $ln\ T$ are identical for large and small farms. A problem arises at this point. Our formulation of the *UOP* profit and labor demand functions is in terms of normalized prices. However, in our empirical application normalized prices are not available since the data on prices of output are rather poor. Fortunately, we note that (17) may be rewritten

$$\begin{aligned} ln\ \Pi_a &= ln\ \Pi' - ln\ p \\ &= ln\ A_*^S + \overset{*}{\delta}_L D_L + \overset{*}{\alpha}_1\ ln\ w' - \overset{*}{\alpha}_1\ ln\ p \\ &\quad + \overset{*}{\beta}_1\ ln\ K + \overset{*}{\beta}_2\ ln\ T \end{aligned}$$

or

$$\begin{aligned} ln\ \Pi' &= ln\ A_*^S + \overset{*}{\delta}_L D_L + (1 - \overset{*}{\alpha})\ ln\ p \\ &\quad + \overset{*}{\alpha}_1\ ln\ w' + \overset{*}{\beta}_1\ ln\ K + \overset{*}{\beta}_2\ ln\ T \end{aligned}$$

where Π' is actual money profit in rupees, w' is the money wage rate in rupees per day, and p is the price of the output in rupees. If the prices of outputs differ only across states, then one can insert state dummy variables to capture the effect of differences due to $(1-\alpha_1^*)\ ln\ p$. This also allows interstate differences in the technical efficiency parameters. Observe that (18) holds independently of the price of output

$$-\frac{wL}{\Pi_a} = -\frac{w'L}{\Pi'} = \alpha_1^{*L} D_L + \alpha_1^{*S} D_S$$

Hence our final estimating equations consist of

$$\text{(19)}\quad \begin{aligned} ln\ \Pi' &= ln\ A_*^S + \overset{*}{\delta}_L D_L + \sum_{i=1}^{4} \overset{*}{\delta}_i D_i \\ &\quad + \overset{*}{\alpha}_1\ ln\ w' + \overset{*}{\beta}_1\ ln\ K \\ &\quad + \overset{*}{\beta}_2\ ln\ T \end{aligned}$$

and

$$\text{(20)}\quad -\frac{w'L}{\Pi'} = \alpha_1^{*L} D_L + \alpha_1^{*S} D_S$$

where Π' is farm profit in rupees and w' is money wage rate.

A remark about the stochastic specification of the model is appropriate at this point. Not much is known about how disturbance terms in general should be introduced into economic relationships although Irving Hoch (1955, 1958) and Mundlak and Hoch and subsequently Arnold Zellner, Jan Kmenta and Jacques Drèze have proposed one possible assumption that is workable in the Cobb-Douglas case. Here we follow the usual, and admittedly *ad hoc* practice of assuming an additive error with zero expectation and finite variance for each of the equations (19) and (20).[12] For the same farm, the covariance of the errors of the two equations is permitted to be nonzero. However, the covariances of the errors of either equation corresponding to different farms are assumed to be identically zero.

Given this specification of errors, it is immediately apparent that Zellner's method provides an asymptotically efficient method of estimation. Moreover, the efficiency of estimation can be increased by imposing known constraints on the coefficients in the equations.

Six statistical hypotheses are tested successively on the data. The results of these tests are reported in Table 1. They are:

(i) Equal Relative Economic Efficiency

$$H_0: \overset{*}{\delta}_L = 0$$

i.e., $A_*^L = A_*^S$ or $ln\ (A_*^L/A_*^S) = 0$. This hypothesis is rejected at the 10 percent significance level. Hence we conclude that small farms are relatively more economic-efficient than large farms.

(ii) Equal Relative Price Efficiency

$$H_0: \alpha_1^{*L} = \alpha_1^{*S}$$

This is not rejected at the 10 percent level. Hence we conclude that small and large

[12] Marc Nerlove, in his pioneering study of cost functions, derives an additive error to the natural logarithm of the cost function. We can do the same here for our profit function, using the same assumptions as his. The additive error in the second equation may arise from different abilities to maximize profits or divergence between expected and realized prices.

TABLE 1—STATISTICAL HYPOTHESES TESTED

Maintained Hypothesis	Tested Hypothesis H_0	Computed F	Critical $F_{0.10}$
	i) $\delta_L^* = 0$	$F(1, 57) = 3.96$	$F(1, 57) = 2.79$
	ii) $\alpha_1^{*L} = \alpha_1^{*S}$	$F(1, 57) = 1.93$	$F(1, 57) = 2.79$
	iii) $\delta_L^* = 0$, $\alpha_1^{*L} = \alpha_1^{*S}$	$F(2, 57) = 4.19$	$F(2, 57) = 2.39$
$\alpha_1^{*L} = \alpha_1^{*S}$	iv) $\alpha_1^{*S} = \alpha_1^{*}$	$F(1, 58) = 0.09$	$F(1, 58) = 2.79$
$\alpha_1^{*L} = \alpha_1^{*S}$	v) $\alpha_1^{*L} = \alpha_1^{*}$	$F(1, 58) = 0.09$	$F(1, 58) = 2.79$
$\left.\begin{array}{l}\alpha_1^{*L} = \alpha_1^{*} \\ \alpha_1^{*S} = \alpha_1^{*}\end{array}\right\}$	vi) $\beta_1^* + \beta_2^* = 1$	$F(1, 59) = 2.01$	$F(1, 59) = 2.79$

Note: For definition of variable see Table 2.
Source: Table 2.

farms do not have different price-efficiency parameters, k_1^i, i.e., they both succeed to the same degree in maximizing profits.

(iii) Equal Relative Technical *and* Price Efficiency

$$H_0: \delta_L^* = 0$$
$$\alpha_1^{*L} = \alpha_1^{*S}$$

This hypothesis is rejected. This actually can be anticipated in view of the rejection of hyphothesis (i).

(iv) Absolute Price Efficiency of Small Farms

$$H_0: \alpha_1^{*S} = \alpha_1^{*}$$

Maintaining the hypothesis of equal price efficiency in (ii), we test for absolute price efficiency of small farms. This hypothesis is not rejected at the 10 percent level, implying that the small farms have maximized profits, i.e., $k_1^S = 1$.

(v) Absolute Price Efficiency of Large Farms

$$H_0: \alpha_1^{*L} = \alpha_1^{*}$$

Also maintaining the hypothesis of equal relative price efficiency of (ii) we test for the absolute price efficiency of large farms. This hypothesis also cannot be rejected at the 10 percent level. This means that large farms have also maximized profits, i.e., $k_1^L = 1$. One could have expected this finding from the results of (ii) and (iv).

(vi) Constant Returns to Scale

$$H_0: \beta_1^* + \beta_2^* = 1$$

Maintaining the hypotheses (iv) and (v) we test for constant returns to scale to all factors of production. Again, this hypothesis cannot be rejected.[13]

Once a hypothesis is not rejected, we proceed to obtain estimates of parameters incorporating the hypothesis. The estimation results are tabulated in Table 2. As predicted by economic theory, the profit function is decreasing and convex in wage rate, and increasing in land. The wrong sign of the coefficient of capital can be attributed to a misspecification of this variable. Defining capital as interest charges on fixed assets per farm involves the implicit assumption that the service flow of capital, which is the appropriate input in production analysis, is proportional to the capital stock. As shown

[13] At this point compare the discussion of A. M. Khusro.

TABLE 2—JOINT ESTIMATION OF COBB-DOUGLAS PROFIT FUNCTION AND LABOR DEMAND FUNCTION

Parameter	Single Equation Ordinary Least Squares	Unrestricted	Zellner's Method with Restrictions: 1 Restriction $\alpha_1^{*L}=\alpha_1^{*S}$	2 Restrictions $\alpha_1^{*L}=\alpha_1^*$ $\alpha_1^{*S}=\alpha_1^*$	3 Restrictions $\alpha_1^{*L}=\alpha_1^*$ $\beta_1^*+\beta_2^*=1$ $\alpha_1^{*S}=\alpha_1^*$
		UOP Profit Function			
$ln\ A_*^S$	4.582	4.132	4.203	4.267	4.576
	(.470)	(.413)	(.416)	(.289)	(.202)
δ_L^*	−.567	−.416	−.509	−.496	−.318
	(.217)	(.200)	(.192)	(.186)	(.148)
δ_1^*	1.614	.991	1.208	.975	1.151
	(.471)	(.412)	(.416)	(.399)	(.388)
δ_2^*	−1.359	−.654	−.696	−.905	−.596
	(1.093)	(.955)	(.964)	(.568)	(.535)
δ_3^*	−.588	−.176	−.200	−.259	−.260
	(.416)	(.364)	(.367)	(.224)	(.229)
δ_4^*	.296	.239	.243	.124	.375
	(.613)	(.536)	(.541)	(.447)	(.419)
α_1^*	−2.141	−.904	−.977	−1.140	−1.097
	(1.029)	(.900)	(.908)	(.373)	(.374)
β_1^*	−.588	−.324	−.339	−.308	−.459
	(.235)	(.206)	(.207)	(.202)	(.175)
β_2^*	1.797	1.479	1.498	1.467	1.459
	(.200)	(.175)	(.176)	(.170)	(.175)
		Labor Demand Function			
	−.661	−.661	−1.209	−1.140	−1.097
α_1^{*L}	(.549)	(.549)	(.412)	(.373)	(.374)
	−1.826	−1.826	−1.209	−1.140	−1.097
α_1^{*S}	(.582)	(.582)	(.412)	(.373)	(.374)

Notes: Numbers in parentheses are asymptotic standard errors. The estimating equations are

$$ln\,\Pi' = ln\,A_*^S + \overset{*}{\delta}_L D_L + \sum_{i=1}^{4} \overset{*}{\delta}_i D_i + \overset{*}{\alpha}_1\, ln\, w' + \overset{*}{\beta}_1\, ln\, K + \overset{*}{\beta}_2\, ln\, T$$

$$-\frac{w'L}{\Pi'} = \overset{*L}{\alpha_1} D_L + \overset{*S}{\alpha_1} D_S$$

where

Π' = profit including interest on fixed capital and land rent
w' = money wage rate
D_L = a size dummy variable taking the value of one for farms with a physical area greater than ten acres and zero otherwise
D_S = a size dummy variable taking the value of one for farms with a physical area less than ten acres and zero otherwise.
D_i = regional dummy variable with D_1, D_2, D_3, D_4 taking the value of one for only West Bengal, Madras, Madhya Pradesh and Uttar Pradesh, respectively.
L = labor in days per year per farm
K = interest on fixed capital per farm
T = cultivable land in acres per farm

Source: The Studies, Delhi, 1957–62. Reports for the year 1955–1965: Madras, Punjab, Uttar Pradesh, West Bengal; Report for the year 1956–1957: Madhya Pradesh.

elsewhere (Yotopoulos (1967a, b)), this measure of capital leads to biased coefficients.

From our estimated parameters, we can compute, by (9) and (11), indirect estimates of the production elasticities of labor, capital and land as well as those of the technical-efficiency and price-efficiency parameters of the large and small farms. These are tabulated in Table 3. We note that the production elasticities thus estimated appear generally large by comparison with estimates of directly fitted Cobb-Douglas agricultural production functions reported in other studies. The obvious reason for this discrepancy

TABLE 3—INDIRECT ESTIMATES OF THE INPUT ELASTICITIES OF THE PRODUCTION FUNCTION AND THE EFFICIENCY PARAMETERS[a]

Parameter		1 Restriction $\alpha_1^{*L} = \alpha_1^{*S}$	2 Restrictions $\alpha_1^{*L} = \alpha_1^{*}$ $\alpha_1^{*S} = \alpha_1^{*}$	3 Restrictions $\alpha_1^{*L} = \alpha_1^{*}$ $\alpha_1^{*S} = \alpha_1^{*}$ $\beta_1^{*} + \beta_2^{*} = 1$
Labor	α_1	.494	.533	.523
Capital	β_1	−.172	−.144	−.219
Land	β_2	.758	.686	.696
Sum of Elasticities ($\alpha_1 + \beta_1 + \beta_2$)		1.080	1.075	1.000
Large Farms	A^L	14.421	11.612	15.228
	k_1^L	0.903	1.000	1.000
Small Farms	A^S	18.658	14.639	17.723
	k_1^S	0.903	1.000	1.000

Note: For definition of variables, see Table 2.

[a] The indirect estimates of the coefficients of the production function are computed from the identities in (9). The efficiency parameters are computed from (9) and (11).

is the constraint of constant returns to scale combined with the negative coefficient of capital.[14] Conceptually, however, the indirect estimates of production coefficients are statistically consistent, as opposed to those obtained directly from the production function by ordinary least squares, which are in general inconsistent because of the existence of simultaneous equations bias. By examining the technical efficiency component for small and large farms in Table 3 we find that the former are approximately 20 percent more efficient than the latter.

IV. Summary and Conclusions

An operational model to measure separately relative technical efficiency and price efficiency between (groups of) firms has been developed and applied to Indian agriculture. Needless to say, the usefulness of the approach here is not restricted to agriculture; neither is it specific for comparing large and small farms.

The conclusion of the test of relative economic efficiency is in favor of the small farms, i.e., farms of less than ten acres, in agreement with our previous result in our 1971 article. Given the fixed inputs (land and fixed capital), and within the ranges of the observed prices of output and variable inputs (labor), the small farms of the sample of *The Studies* have higher actual profits. The relative economic efficiency of the small farms, however, is not due to superior price efficiency. In fact, we have found that both large and small farms are price-efficient.[15] It may, after all, be true that the marginal calculations which the market place requires of the farmer are relatively easy to make. More importantly, we have established the superior technical efficiency of the small farms.[16] We have also found that there exist constant returns to scale in Indian agriculture. Hence one cannot argue for a consolidation of small farms on the grounds of economies of scale.

These findings imply that in agriculture the direct supervisory and leadership role of

[14] The negative coefficient of capital has also appeared in other application with the same data. See Lau and Yotopoulos (1971) and Yotopoulos, Lau, and K. Somel.

[15] This finding is inconsistent with Amartya Sen's (1964, 1966) hypothesis that large farms maximize profits while small farms maximize utility.

[16] One possible explanation that has been advanced is an inverse proportional relationship between farm size and intrinsic fertility.

the owner-manager may be crucial for attaining high levels of economic efficiency. More studies are needed to further explain the so-called "unmeasureable factors" which apparently cause the difference in technical efficiency.

REFERENCES

M. J. Farrell, "The Measurement of Productive Efficiency," *J. Royal Statist. Soc.*, 1957, *120*, 253–82.

I. Hoch, "Estimation of Production Function Parameters and Testing for Efficiency," *Econometrica*, July 1955, *23*, 325–26.

———, "Simultaneous Equation Bias in the Context of the Cobb-Douglas Production Function," *Econometrica*, Oct. 1958, *26*, 556–78.

W. D. Hopper, "Allocation Efficiency in a Traditional Indian Agriculture," *J. Farm Econ.*, Aug. 1965, *47*, 611–24.

A. M. Khusro, "Returns to Scale in Indian Agriculture," *Indian J. Agric. Econ.*, Oct. 1964, *19*, 51–80.

L. J. Lau, "Applications of Profit Functions," in D. L. McFadden, ed., *An Econometric Approach to Production Theory*, Amsterdam 1972, forthcoming.

——— and P. A. Yotopoulos, "A Test for Relative Efficiency and an Application to Indian Agriculture," *Amer. Econ. Rev.*, Mar. 1971, *61*, 94–109.

——— and ———, "Profit, Supply, and Factor Demand Functions with Application to Indian Agriculture," *Amer. J. Agric. Econ.*, Feb. 1972, *54*, 11–18.

D. L. McFadden, "Cost, Revenue, and Profit Functions," in D. L. McFadden, ed., *An Econometric Approach to Production Theory*, Amsterdam 1972, forthcoming.

Y. Mundlak, "Empirical Production Function Free of Management Bias," *J. Farm Econ.*, Feb. 1961, *43*, 44–56.

——— and I. Hoch, "Consequences of Alternative Specifications in Estimation of Cobb-Douglas Production Functions," *Econometrica*, Oct. 1965, *33*, 814–28.

M. Nerlove, "Returns to Scale in Electricity Supply," in C. F. Christ et al., eds., *Measurement in Economics: Studies in Mathematical Economics and Econometrics in Memory of Yehuda Grunfeld*, Stanford 1960, 167–98.

T. W. Schultz, *Transforming Traditional Agriculture*, New Haven 1964.

W. D. Seitz, "The Measurement of Efficiency Relative to a Frontier Production Function," *Amer. J. Agric. Econ.*, Nov. 1970, *52*, 505–11.

A. K. Sen, "Size of Holdings and Productivity," *Econ. Weekly*, Feb. 1964, *16*, 323–26.

———, "Peasants and Dualism With or Without Surplus Labor," *J. Polit. Econ.*, Oct. 1966, *74*, 425–50.

R. M. Solow, "Technical Progress and the Aggregate Production Function," *Rev. Econ. Statist.*, Aug. 1957, *39*, 312–20.

C. P. Timmer, "On Measuring Technical Efficiency," *Food Res. Inst. Stud. in Agric. Econ., Trade, Develop.*, 1970, *9*, 99–171.

J. Wise and P. A. Yotopoulos, "The Empirical Content of Rationality: A Test for a Less Developed Economy," *J. Polit. Econ.*, Nov. 1969, *77*, 976–1004.

P. A. Yotopoulos, (1967a) *Allocative Efficiency in Economic Development: A Cross Section Analysis of Epirus Farming*, Athens 1967.

———, (1967b) "From Stock to Flow Capital Inputs for Agricultural Production Functions: A Microanalytic Approach," *J. Farm Econ.*, May 1967, *49*, 476–91.

———, "On the Efficiency of Resource Utilization in Subsistence Agriculture," *Food Res. Inst. Stud. in Agric. Econ., Trade, Develop.*, 1968, *8*, 125–35.

———, L. J. Lau, and K. Somel, "Labor Intensity and Relative Efficiency in Indian Agriculture," *Food Res. Inst. Stud. in Agric. Econ., Trade Develop.*, 1970, *9*, 43–55.

A. Zellner, "An Efficient Method for Estimating Seemingly Unrelated Regressions and Tests for Aigregation Bias," *J. Amer. Statist. Ass.*, June 1962, *57*, 348–68.

———, J. Kmenta, and J. Drèze, "Specification and Estimation of Cobb-Douglas Production Function Models," *Econometrica*, Oct. 1966, *34*, 784–95.

Government of India, Ministry of Food and Agriculture, *Studies in the Economics of Farm Management*, Delhi 1957–62, reports for the year 1955–56: Madras, Punjab, Uttar Pradesh, West Bengal; report for the year 1956–57: Madhya Pradesh.

Welfare-Maximizing Price and Output with Stochastic Demand: Comment

By Michael L. Visscher*

In their model of a public utility firm which faces constant returns to scale and stochastic demand, Gardner Brown, Jr. and M. Bruce Johnson argued that "... the optimal price will always be lower and, with linear demand, the optimal output will generally be higher than their counterparts in the riskless model of traditional theory, regardless of the manner in which the disturbance term enters the analysis" (p. 119). Although their formal analysis is faultless, they are in error in attributing primary responsibility for their conclusions to the random element in the demand schedule. Their results can be shown to depend instead on their assumption of an unrealistic rationing system; when quantity demanded at the predetermined price exceeds capacity, they assume that available output will be allocated to those with the highest demand prices. Hence, the Brown and Johnson (B-J) approach assumes away the unique role of a market-clearing price in achieving an efficient rationing of the available product. It is the purpose of this note to show that if alternative assumptions are made concerning the way in which output is allocated when excess demand exists, different conclusions are reached about the nature of welfare-maximizing price and investment rules when demand is stochastic.

I. The Brown and Johnson Model

In one of the B-J models, a random disturbance term enters the demand function additively: $D=X(P)+u$, where D is actual quantity demanded, P is price, and u is a random variable which is distributed symmetrically about zero with continuous probability density function $f(u)$ and cumulative probability function $F(u)$. The public utility firm has the long-run cost function $C(Q)=(b+\beta)Q$, where Q is quantity produced, b is marginal operating cost, β is marginal capacity cost, and both b and β are constant. Before the actual value of u is known, the firm must choose a price $\overline{P}$ and a capacity Z to maximize a welfare function equal to expected consumers' surplus: $W=E$(willingness to pay) $-E$(average variable cost·sales) $-E$(capacity cost).

In order for the B-J model to be appropriate, the producer must be able to distinguish costlessly between claimants (i.e., all those willing to pay $\overline{P}$) and serve only those with the highest willingness to pay. Assuming this rationing procedure is possible, differentiating W with respect to the two choice variables, $\overline{P}$ and Z, yields marginal conditions

$$(1)\quad \frac{\partial W}{\partial \overline{P}} = \overline{P}X'(\overline{P})F[Z-X(\overline{P})] - bX'(\overline{P})F[Z-X(\overline{P})]=0$$

$$(2)\quad \frac{\partial W}{\partial Z} = \int_{Z-X(P)}^{\infty} f(u)\cdot[X^{-1}(Z-u)-b]du-\beta=0$$

From these conditions, Brown and Johnson conclude that the optimal price is b and that optimal capacity is greater than $X(b+\beta)$, the optimal capacity in a model of riskless deterministic demand.[1]

The B-J conclusion must, however, be qualified because (a) the optimal values of price and capacity depend critically upon which claimants are actually served, and (b) in the absence of a market-clearing price, the B-J rationing procedure simply is unavailable to the producer. When other assumptions are made about who is served, different solutions follow. Two alternative rationing procedures will be discussed.

* Graduate student, University of Virginia. This paper has benefited significantly from the comments of Roger Sherman and an anonymous referee. Remaining errors are my own.

[1] See Oliver Williamson, p. 812, for development of the riskless model.

II. Serving First Those with the Lowest Willingness to Pay

The B-J assumption can be depicted graphically in Figure 1. When a value of the error term, u is sufficiently high to result in excess demand at the previously chosen values of $\bar{P}$ and Z, available output is sold only to those claimants with the highest willingness to pay. Then the sum of consumers' surplus and total revenue is represented by the area $ABZO$ in Figure 1.

Suppose instead that service is offered first to those claimants with the least willingness to pay. Such an assumption may be relevant if, for instance, distribution is on a first-come, first-served basis, and those with the lowest opportunity costs on their time (and therefore most willing to pay by waiting) are also those with the lowest demand prices. Then the sum of consumers' surplus and total revenue in times of excess demand could be depicted as area $BCDE$ in Figure 2.

Analysis of this model can proceed in a manner similar to that of Brown and Johnson. The objective function will remain: $W = E$(willingness to pay) $- E$(average variable costs·sales) $- E$(capacity cost). If demand always could be satisfied, E(willingness to pay) would be the expected value of all possible values of consumers' surplus plus total revenue at the price chosen. But

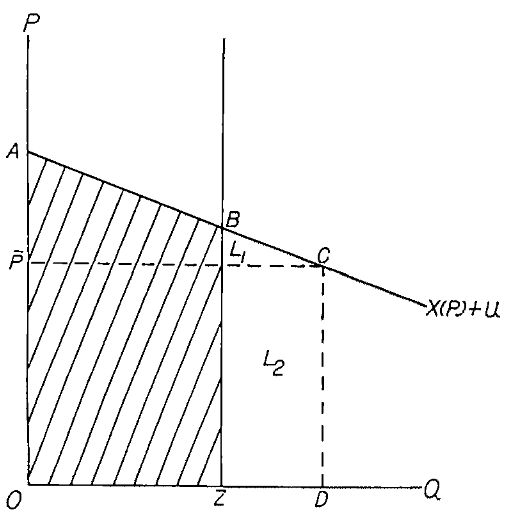


FIGURE 1

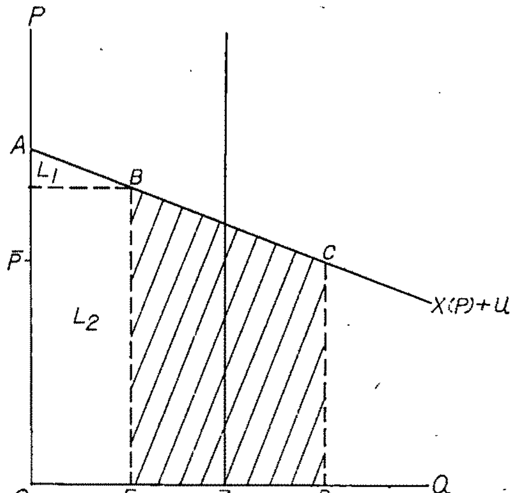


FIGURE 2

at some values of u, the demand cannot be satisfied. Such a value of u is depicted by the position of the demand function in Figures 1 and 2. If this high value of u manifests itself and those with the lowest willingness to pay are served first, (L_1+L_2) in Figure 2 must be subtracted from area $ACDO$ to arrive at total willingness to pay. The expected value of (L_1+L_2) in Figure 2 is:

$$(3) \quad E(L_1 + L_2) = \int_{Z-X(P)}^{\infty} f(u) \cdot \left\{ \int_{X^{-1}(-u)-X^{-1}(Z-u)+P}^{X^{-1}(-u)} [X(P) + u]dP + [X^{-1}(-u) - X^{-1}(Z-u) + \bar{P}] \cdot [X(\bar{P}) + u - Z] \right\} du$$

On subtracting this expected value, the welfare function becomes that shown in equation (4). (See following page.)

To maximize this objective function, we differentiate with respect to the two choice variables and set each differentiation equal to zero as shown in equations (5) and (6). Adding these two equations yields $\bar{P} = b + \beta$. Thus, in a model in which available output is rationed to those claimants with the smallest demand prices, optimal price is equal to

$$
(4) \quad \begin{aligned}
W = &\int_{-\infty}^{\infty} f(u) \int_{P}^{X^{-1}(-u)} [X(P) + u] dP du + \overline{P}X(\overline{P}) - \int_{Z-X(P)}^{\infty} f(u) \\
&\cdot \int_{X^{-1}(-u)-X^{-1}(Z-u)+P}^{X^{-1}(-u)} [X(P) + u] dP du \\
&- \int_{Z-X(P)}^{\infty} f(u)[X^{-1}(-u) - X^{-1}(Z-u) + \overline{P}][X(\overline{P}) + u - Z] du \\
&- b\left\{X(\overline{P}) - \int_{Z-X(P)}^{\infty} uf(u)du + [Z - X(\overline{P})] \int_{Z-X(P)}^{\infty} f(u)du\right\} - \beta Z
\end{aligned}
$$

$$
(5) \quad \frac{\partial W}{\partial P} = (\overline{P} - b)F[Z - X(\overline{P})] - \int_{Z-X(P)}^{\infty} f(u)[X^{-1}(-u) - X^{-1}(Z-u)]du = 0
$$

$$
(6) \quad \frac{\partial W}{\partial Z} = (\overline{P} - b)\left\{1 - F[Z - X(\overline{P})]\right\} + \int_{Z-X(P)}^{\infty} f(u)[X^{-1}(-u) - X^{-1}(Z-u)]du - \beta = 0
$$

$$
(7) \quad \left.\frac{\partial W}{\partial Z}\right|_{\substack{P=b+\beta \\ Z=X(b+\beta)}} = \int_{0}^{\infty} f(u)[X^{-1}(-u)]du - \int_{0}^{\infty} f(u)\left\{X^{-1}[X(b+\beta) - u]\right\} du - \frac{\beta}{2}
$$

long-run marginal cost just as in traditional riskless theory, and not short-run marginal cost as required under the rationing assumption made by Brown and Johnson.

Optimal capacity choice in the Figure 2 model is likewise different from that in the B-J model. Smaller values of Z (smaller even than riskless Z) are possible.[2]

If, when evaluated at

$$\overline{P} = b + \beta$$

and

$$Z = X(b+\beta),$$

$\partial W/\partial Z$ is negative, then optimal capacity is less than $X(b+\beta)$, for then increments to capacity at that output decrease the objective function. Yet if Z can be less than $X(b+\beta)$, it can surely be less Brown and Johnson Z, for that could never be less than $X(b+\beta)$. From equation (7), we can assert that optimal capacities less than those of the riskless model are indeed feasible.[3]

[2] In general, Brown and Johnson's conjecture, p. 123, fn. 8, on this point does not appear to hold.

[3] The same is true for a profit-maximizing monopolist under risk. See Edwin Mills, p. 93.

$X^{-1}(-u)$ is the price intercept of the observed demand function, and $X^{-1}(X(b+\beta)-u)$ is the price that would just clear the market for the observed value of u (Figure 3). Therefore, if demand is linear,

$$\left.\frac{\partial W}{\partial Z}\right|_{\substack{P=b+\beta \\ Z=X(b+\beta)}}$$

tends to decrease, *ceteris paribus*, if:

a) the demand function becomes more elastic at $P = b+\beta$ (for then the sum of the first two terms in (7) is less),

b) the short-run marginal cost, b, increases (for then the second term in (7) is less), or

c) the marginal capacity cost, β, increases (for then the last term in (7) is less). It is conceivable that any of these events might be sufficient to make optimal Z less than $X(b+\beta)$.

These solutions for $\overline{P}$ and Z can be squared with intuition. The penalty for

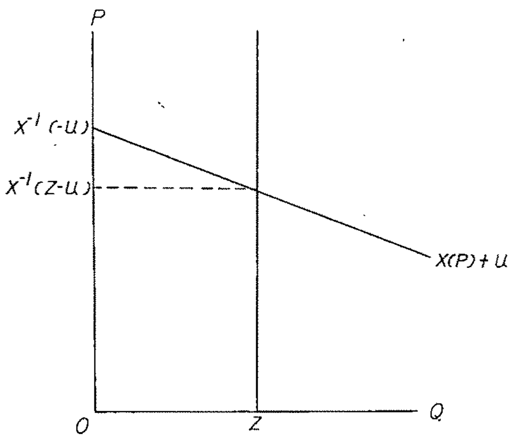


FIGURE 3

excess demand is that those with the highest willingness to pay will be denied service. To avoid this penalty, excess demand could be made less frequent by increasing both price and capacity. But at any given price, additions to capacity will be used to satisfy those who value the good relatively little. So, if the linear demand curve is sufficiently elastic at the higher price, or if either b or β are sufficiently high, then those that would benefit from additional capacity do not value the good enough to warrant the outlays extra capacity would entail. Output should then be restricted. Excess demand is made more frequent, but consumer surplus is conserved; those who are turned away value the service little relative to the cost of production.

III. Random Rationing of Available Capacity

Perhaps a more tenable assumption than either Brown and Johnson's or "least-valued served-first" is that order of service and willingness to pay are completely uncorrelated. If $X(\overline{P})+u$ units are demanded at $\overline{P}$ and only $Z<X(\overline{P})+u$ are available, let those Z units be distributed randomly among those claimants willing to pay $\overline{P}$. The expected value of the willingness to pay in periods of excess demand would then be $Z/[X(\overline{P})+u]$ of that which would be realized if the entire quantity-demanded could be met at the price $\overline{P}$. Therefore, we can calculate E(willingness to pay) by subtracting $E(L_1+L_2)$ from the expected value of area $AEDO$ (Figure 4) for those values of u large enough to cause excess demand at the $\overline{P}$ and Z chosen. Since $E(L_1+L_2)$ is given in equation (8), our objective function is equation (9). Maximizing the objective function with respect to $\overline{P}$ and Z, we obtain equations (10) and (11). The third term on the right-hand side of equation (11) is the expected value of area $AE\overline{P}$ divided by length $\overline{P}E$ (Figure 4) for all ($\overline{P}$, Z, u) combinations where excess demand occurs. The fourth term on the right-hand side is the expected value of area $AB\overline{P}$ divided by length $\overline{P}E$ for ($\overline{P}$, Z, u) combinations where excess demand occurs. This term must be less than the third term because realized consumers' surplus in times of excess demand is always less than that had all demand been met. Thus, we have $b<\overline{P}<b+\beta$.

Here, too, optimal Z can be less than that which was optimal in the B–J and riskless models. Let $\overline{P}=b+\alpha\beta$, where $0<\alpha<1$, and $Z=X(b+\beta)$ in $\partial W/\partial Z$. If $\partial W/\partial Z$ evaluated at that $\overline{P}$ and Z can be less than zero, optimal capacity can lie below $X(b+\beta)$, the riskless optimal capacity. See equation (12), page 228.

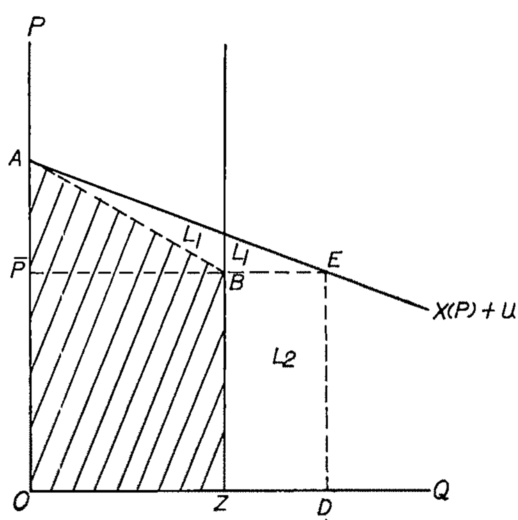


FIGURE 4

$$E(L_1 + L_2) = \int_{Z-X(P)}^{\infty} f(u)\left[1 - \left(\frac{Z}{X(\overline{P}) + u}\right)\right]\int_{P}^{X^{-1}(-u)} [X(P) + u]dPdu \tag{8}$$

$$+ \int_{Z-X(P)}^{\infty} f(u)[\overline{P}][X(\overline{P}) + u - Z]du$$

$$W = \int_{-\infty}^{\infty} f(u)\int_{P}^{X^{-1}(-u)} [X(P) + u]dPdu + \overline{P}X(\overline{P}) - \int_{Z-X(P)}^{\infty} f(u) \tag{9}$$

$$\cdot\left[1 - \left(\frac{Z}{X(\overline{P}) + u}\right)\right]\int_{P}^{X^{-1}(-u)} [X(P) + u]dPdu$$

$$- \int_{Z-X(P)}^{\infty} f(u)[\overline{P}][X(\overline{P}) + u - Z]du - bX(\overline{P}) + b$$

$$\cdot \int_{Z-X(P)}^{\infty} uf(u)du - b[Z - X(\overline{P})]\int_{Z-X(P)}^{\infty} f(u)du - \beta Z$$

$$\overline{P} = b + \frac{1}{F[Z - X(\overline{P})]}\int_{Z-X(P)}^{\infty} \frac{Zf(u)}{[X(\overline{P}) + u]^2}\int_{P}^{X^{-1}(-u)} [X(P) + u]dPdu > b \tag{10}$$

$$\overline{P} = b + \beta - \int_{Z-X(P)}^{\infty} \frac{f(u)}{[X(\overline{P}) + u]}\int_{P}^{X^{-1}(-u)} [X(P) + u]dPdu \tag{11}$$

$$+ \int_{Z-X(P)}^{\infty} \frac{Zf(u)}{[X(\overline{P}) + u]^2}\int_{P}^{X^{-1}(-u)} [X(P) + u]dPdu < b + \beta$$

$$\left.\frac{\partial W}{\partial Z}\right|_{\substack{P=b+\alpha\beta \\ Z=X(b+\beta)}} = \alpha\beta\int_{X(b+\beta)-X(b+\alpha\beta)}^{\infty} f(u)du - \beta + \int_{X(b+\beta)-X(b+\alpha\beta)}^{\infty} \frac{f(u)}{[X(b + \alpha\beta) + u]} \tag{12}$$

$$\cdot \int_{b+\alpha\beta}^{X^{-1}(-u)} [X(P) + u]dPdu$$

$$\left.\frac{\partial W}{\partial Z}\right|_{\substack{P=b+\alpha\beta \\ Z=X(b+\beta)}} = \alpha\beta\int_{X(b+\beta)-X(b+\alpha\beta)}^{\infty} f(u)du - \beta + \frac{1}{2}\int_{X(b+\beta)-X(b+\alpha\beta)}^{\infty} f(u)[X^{-1}(-u) - (b + \alpha\beta)]du \tag{13}$$

$$= \frac{1}{2}\left\{\int_{X(b+\beta)-X(b+\alpha\beta)}^{\infty} f(u)[X^{-1}(-u)]du\right.$$

$$\left. + (\alpha\beta - b)\int_{X(b+\beta)-X(b+\alpha\beta)}^{\infty} f(u)du - 2\beta\right\}$$

If $X(P)$ is linear (as Brown and Johnson also assume), then $\partial W/\partial Z$ is given by equation (13), page 228. The value of $\partial W/\partial Z$ may be less than zero, as in the case of the "least-valued served-first" model, and is more likely to be negative the more elastic is the demand function at $\bar{P}$, the larger is β, and the larger is b, *ceteris paribus.*

IV. Conclusion

Were all transactions costless, we would not need to concern ourselves for the sake of efficiency with the manner in which the producer allocates service among potential claimants when excess demand exists. An implicit market-clearing price is then established regardless of the price set by the producer because of the possibility of trades among claimants. For instance, if the producer chose the B–J price of b, those consumers with the highest demand prices would bid away service from those with the lowest demand prices by offering appropriate side payments. The result would be efficient (in the static sense) since only those with the highest willingness to pay would be served. But such transactions are prohibitively costly, so when demand is uncertain and more service may be demanded than can be supplied at the announced price, the optimal values of price and capacity will depend on how available capacity is initially rationed.

The relation between the rationing assumption which governs who is served and the optimal values of price and capacity can be summarized in the following table:

Model	Price	Capacity
Riskless (Deterministic) Demand	$b+\beta$	$X(b+\beta)$
Highest Demand Prices Served First (B-J)	b	$>X(b+\beta)$
Lowest Demand Prices Served First	$b+\beta$	$X(b+\beta)$ depending on cost and demand parameters
Capacity Distributed Randomly Among All Willing to Buy at $\bar{P}$	$b+\alpha\beta$ $0<\alpha<1$	$X(b+\beta)$ depending on cost and demand parameters

For the B–J results to hold, the producer must insure that only those with the highest willingness to pay are served. But, except by price, no feasible means of distinguishing among claimants exists. Consequently, optimal capacity will be different from that defined by Brown and Johnson, and optimal price will be nearer to that of the riskless model.

REFERENCES

G. Brown Jr. and M. B. Johnson, "Public Utility Pricing and Output Under Risk," *Amer. Econ. Rev.*, Mar. 1969, *59*, 119–28.

E. Mills, *Price, Output, and Inventory Policy,* New York 1962.

O E. Williamson, "Peak-Load Pricing and Optimal Capacity Under Indivisibility Constraints," *Amer. Econ. Rev.*, Sept. 1966, *56*, 810–27.

Welfare-Maximizing Price and Output with Stochastic Demand: Reply

By Gardner Brown, Jr. and M. Bruce Johnson*

We concluded in our 1969 article that when a public monopoly must choose both price and capacity output before it knows the position of the stochastic demand curve, optimal price should equal short-run marginal cost and optimal capacity output should, in general, exceed riskless capacity output. Michael Visscher claims our results depend on an unrealistic rationing system. He proposes two alternative rationing schemes: distribution of the product to those customers with the lowest (rather than the highest) willingness to pay, or distribution to customers on a random basis without regard for willingness to pay. His intuition is correct; different assumptions lead to different conclusions.

Our analysis supressed the costs of identifying those customers with the highest willingness to pay. Visscher challenges this assumption but de facto adopts it himself when he supposes that the product is offered to those claimants with the least willingness to pay. He asserts that if distribution is on a first-come, first-served basis and if consumers "with the lowest opportunity costs on their time (and therefore most willing to pay by waiting) are also those with the lowest demand prices," his rationing scheme is relevant. The second assertion is gratuitous.

There are neither theoretical nor empirical grounds to support the view that willingness to pay is correlated with opportunity cost; even if one were willing to assume that the value of time is equal to the wage rate, the result would not follow. Moreover, in some cases the decision to consume the product requires commitment of a block of time to the activity before one knows whether or not there will be excess demand. For example, the decision to camp in a public park involves a trip to the park site. Once at the site the decision to enter a queue (if one exists) depends on one's subjective disutility of queuing, not on the money wage rate one would have earned if one had remained at his job site. Consequently, if one is at the entrance to the park, the opportunity cost of one's time in productive activities probably is zero. By implication, one cannot conclude that those customers with the lowest opportunity cost are most likely to stand in line for the product.

In any event, neither our model nor the neoclassical excludes the possibility that those potential consumers with the highest willingness to pay will hire other consumers with low opportunity cost to stand in line for them. More formally, a reservation or a right to use the product may be transferred among customers. The initial distribution of rights is irrelevant since free trade among the participants will redistribute the product to those customers with the highest willingness to pay. Visscher's first model would be relevant as stated only if resale and transfer of rights were prohibited by law or if high customer search and transfer costs existed.

According to our model, optimum price equals short-run marginal cost b. Given that long-run marginal (and average cost) is $b+\beta$, the public monopoly will incur a loss. However, the public authority can recover its fixed cost by selling rights to use the Z^* units of capacity at a price of β each in a futures market. The operation will "break even" since $(b+\beta)\ Z^*$ equals both receipts and expenditures. Speculators will purchase rights at the price of β as long as the market includes one risk neutral trader or, in the vernacular of J. M. Keynes, pp. 142–44, if the futures market does not exhibit normal

* Departments of economics, University of Washington and University of California, Santa Barbara, respectively.

backwardation, i.e., if speculators do not demand a risk premium.[1]

On the average, those customers who desire a product would be willing to pay:

$$\int_{Z-X(p)}^{\infty} f(u)[X^{-1}(Z-u)-b]du=\beta$$

in the futures market and b per unit at the time of delivery. On the average speculators as a group will receive β per unit. This futures market performs its classical function: public authorities are able to hedge against risk and those individuals with a flair for risk management have the opportunity to exercise their talents. Our explicit introduction of a futures market in the section titled "The Way Back," shifted a portion of consumers' surplus (a rectangle whose dimensions are β and Z^*) from the consumers to the producer. Since the producer and consumer components have equal weight in our welfare function, our measure of welfare is unchanged by the introduction of futures markets. Furthermore, our welfare index will display a higher value than the index implicit in Visscher's solution where the product is distributed randomly, or to those individuals with the lowest willingness to pay.

On balance, Visscher's contribution is a constructive one. Transactions and recontracting costs have been troublesome in neoclassical static price theory for some time. However, it is easier to note their existence and speculate on their importance than it is to specify the exact details of a "real time" distribution scheme in a systematic and rigorous way. That task remains to be done both in our model and in the neoclassical competitive model.

[1] Years ago some futures markets for inactively traded commodities did display normal backwardation (see Roger Gray), but contemporary evidence suggests that speculators in 25 of the major commodity futures markets are risk neutral (see Charles Rockwell, pp. 107, 109). The profession remains divided on this issue (see Paul Cootner).

Erratum

The conclusion of Section I-A, page 124 of our 1969 article is correct: $Z^*>\bar{Z}$. However, the argument is imperfect due to the omission of the coefficient β from equation (17). Add β to the left-hand side of (17) and rearrange to give:[2]

$$\delta=-\frac{1}{B}\left[B\beta\int_{-\infty}^{B\beta} f(u)du+\int_{B\beta}^{\infty} uf(u)du\right] \tag{17'}$$

But

$$\int_{B\beta}^{\infty} uf(u)du=-\int_{-\infty}^{B\beta} uf(u)du \quad \text{since } E[u]=0$$

Thus,

$$\delta=-\frac{1}{B}\int_{-\infty}^{B\beta} f(u)[B\beta-u]du \quad \text{after substitution}$$

The limits on the integral are such, that $-\infty<u\leq B\beta$ where $B<0$ and $\beta>0$ by construction. $u\leq B\beta$ implies $[B\beta-u]\geq 0$. Hence the integral is positive and $\delta>0$.

REFERENCES

G. Brown and M. B. Johnson, "Public Utility Pricing and Output Under Risk," *Amer. Econ. Rev.*, Mar. 1969, *59*, 119–28.

P. Cootner, "Speculation and Hedging," *Food Res. Inst. Stud.*, 1967, suppl., *7*, 65–106.

R. W. Gray, "The Characteristic Bias on Some Thin Futures Markets," *Food Res. Inst. Stud.*, Nov. 1960, *1*, 296–312.

J. M. Keynes, *A Treatise on Money*, Vol. II, London 1953.

C. S. Rockwell, "Normal Backwardation Forecasting and the Returns to Commodity Futures Trades," *Food Res. Inst. Stud.*, 1967, suppl., *7*, 107–30.

[2] We are endebted to Susan Wilcox for a suggestion that shortened this proof.

Dependency Rates and Savings Rates: Further Comment

By ARTHUR S. GOLDBERGER*

The empirical results on dependency rates and savings rates reported by Nathaniel Leff (1969) cannot be correct.

For several cross-country samples, Leff estimates pairs of equations of the form

$$(1)\quad y_1 = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 + e_1$$

$$(2)\quad y_2 = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + e_2$$

where

$y_1 = ln\ S/Y = ln$ domestic savings ratio
$y_2 = ln\ S/N = ln$ per capita savings
$x_1 = ln\ Y/N = ln$ per capita income
$x_2 = g =$ growth rate of per capita income
$x_3 = ln\ D_1 = ln$ percentage of population aged 14 or less
$x_4 = ln\ D_2 = ln$ percentage of population aged 65 or more,

and a and b are least-squares regression coefficients, and e_1 and e_2 are least-squares residuals.

As noted by Leff, $S/N \equiv (S/Y)(Y/N)$. Consequently, $y_2 = y_1 + x_1$. Least-squares regression being what it is, a proper computation of (2) should produce

$$(3)\quad y_2 = a_0 + (1 + a_1)x_1 + a_2x_2 + a_3x_3 + a_4x_4 + e_1$$

That is, regressing y_2 on the x should give the same coefficients and the same residuals as occur when y_1 is regressed on the x, except for the coefficient of x_1, which should increase by exactly 1.

Furthermore, if regression coefficients are guaranteed to be equal, their standard errors, and hence their t-ratios, must be equal. If regression coefficients are guaranteed to differ by unity, their standard errors must be equal, and hence their t-ratios must be related by

$$b_1/s_{b_1} = (a_1/s_{a_1})((1 + a_1)/a_1)$$

But the results Leff reports do not satisfy these arithmetic requirements. For example, consider the upper panel of his Table 1, p. 891, which refers to a sample of 47 underdeveloped countries. In the present notation we find:

Explanatory variable $j=$	1	2	3	4
Coefficients:				
a_j	.1292	.0227	−1.2297	−.4455
b_j	1.1167	.0239	−1.3122	−.4469
t-ratios:				
a_j/s_{a_j}	1.8487	2.8079	2.7636	2.1554
b_j/s_{b_j}	16.8355	3.1204	2.9400	2.2783

Something must be wrong.

For the same reasons, the empirical results reported by Kanhaya Gupta cannot be correct. Gupta fits (1) and (2) to 3 subgroups of Leff's sample, using Leff's data. His results also fail to satisfy the arithmetic requirements noted above. For example, consider the first panel of his Table 2, p. 470, which refers to 9 underdeveloped countries. In the present notation we find:

Explanatory variable $j=$	1	2	3	4
Coefficients:				
a_j	.4548	.1263	−.7685	−.6475
b_j	1.6112	.1309	−1.0487	−.5585
t-ratios:				
a_j/s_{a_j}	.3768	.3737	.1481	.9446
b_j/s_{b_j}	1.6008	.4644	.2424	.9770

The discrepancies are somewhat more striking in this subsample.

Perhaps none of the discrepancies noted here is large enough to vitiate the substantive analyses of Leff and Gupta. But it is disturbing that figures, some of which are

* Professor of economics, University of Wisconsin, Madison.

transparently wrong in the first decimal place, are reported to four decimal places.

One's first reaction is to attribute such discrepancies to computational inaccuracies. (Indeed, G. M. Mullet and T. W. Murray recently proposed fitting equation pairs like (1)–(2) to check the accuracy of computer programs.) But in the present case, the explanation lies elsewhere. The basic data constructed by Leff, and used by him and Gupta, were internally inconsistent. Scanning the series, which were kindly provided to me by Leff, one finds that, country by country, there are substantial discrepancies between per capita savings and the product of the savings ratio times per capita income. For example, for the first two countries in the list, we find

S/N	S/Y	Y/N	$(S/Y)(Y/N)$
43.7	.101	585	59.1
20.6	.094	271	25.5

Since $y_2 \neq y_1 + x_1$, the argument which led us to (3) loses its force. But it's difficult to see how an economic analysis of the relationship between savings rates and dependency rates can be based on data which are so internally inconsistent.

REFERENCES

K. L. Gupta, "Dependency Rates and Savings Rates: Comment," *Amer. Econ. Rev.*, June 1971, *61*, 469–71.

N. H. Leff, "Dependency Rates and Savings Rates," *Amer. Econ. Rev.*, Dec. 1969, *59*, 886–96.

———, "Dependency Rates and Savings Rates: Reply," *Amer. Econ. Rev.*, June 1971, *61*, 476–80.

G. M. Mullett and T. W. Murray, "A New Method for Examining Rounding Error in Least-Squares Computer Programs," *J. Amer. Statist. Ass.*, Sept. 1971, *66*, 496–98.

Dependency Rates and Savings Rates: Reply

By NATHANIEL H. LEFF*

Arthur Goldberger's comment is a tempest in a teacup. This is readily verified from the data presented in Table 1. This table reproduces from my original paper the coefficients for paired equations estimated with three different samples in which $ln\ S/N$ and $ln\ S/Y$ were the dependent variables.

As the data of Table 1 indicate, in 7 of the 12 cases, the coefficients are identical through the second decimal place. Furthermore, all the discrepancies are quantitatively very small. In 9 cases, the paired parameter estimates are within 1 percent of each other. In the remaining cases, the differences are of the magnitude of .0825, .0012, and .0011. These are well within the margins of accuracy within which anyone views data generated from the national income accounts of the less developed countries. In light of these considerations, it is difficult to understand the tone of Goldberger's comment.

The reason for the slight discrepancies is straightforward. My observations for S/N and S/Y were computed separately from data supplied by the Statistical Office of the Agency for International Development. With the rounding introduced by the series for N and Y, it is not surprising that the series for S/N is not identically equal to the series which would be obtained by forming S/N from the product of Y/N and S/Y. Similar slight discrepancies have also been reported in other econometric work in which equations for S/N and S/Y have been estimated with separately computed data series.[1]

In any case, nowhere in my paper did I even refer to the precise numerical parameter estimates obtained in my equations. The main conclusions of the paper were that the dependency variables were quantitatively an important determinant of international savings rates, and, further, that introduction of these variables greatly reduced the importance of percapita income, on which some previous discussions had focused. Neither of these conclusions is affected by Goldberger's comment.

REFERENCES

L. Landau, "Determinants of Savings in Latin America," *Project for Quantitative Research on Economic Development,* Harvard Univ. 1966.

* Associate professor, Graduate School of Business, Columbia University.

[1] See, for example, Luis Landau, p. 6, equations (5) and (7).

TABLE 1—PAIRWISE COMPARISON OF PARAMETER ESTIMATES

Sample	$a_1 \approx b_1 - 1$	b_1	$a_2 \approx b_2$		$a_3 \approx b_3$		$a_4 \approx b_4$	
74 Countries	.1596	1.1486	.0254	.0265	−1.3520	−1.3438	−.3990	−.3966
47 Underdeveloped Countries	.1292	1.1167	.0227	.0239	−1.2297	−1.3122	−.4455	−.4469
20 Western Developed Countries	.0035	1.0049	.2589	.2591	−.4324	−.4300	−.4916	−.4914

Source: Data presented on pp. 889–891 of my 1969 article.
Note: The constants of equations (1) and (2) differ by approximately 4.6 (equal the antilog of 100). This shift by a factor of 100 occurs because, as stated on p. 888 of my paper, S/Y is expressed in percentage *points.*
Equation (1): $ln\ S/Y = a_0 + a_1\ ln\ Y/N + a_2\ ln\ g + a_3\ ln\ D_1 + a_4\ ln\ D_2$
Equation (2): $ln\ S/N = b_0 + b_1\ ln\ Y/N + b_2\ ln\ g + b_3\ ln\ D_1 + b_4\ ln\ D_2$

Behavior of the Firm Under Regulatory Constraint: Clarifications

By MOHAMED EL-HODIRI AND AKIRA TAKAYAMA*

In the note by Israel Pressman and Arthur Carol (P-C) relating to Takayama's paper (T) on the Averch-Johnson effect (A-J), it seems useful to clarify the following points.

I. The Shift of the Curve

The major point of criticism on A-J by T, is that (A-J) "... involves an error of confirming the movements along the curve with the shift of the curve" (T, p. 257). This point seems to be obvious from Figure 1 in T, p. 258, but it looks like P-C has a serious objection to this (see p. 211). Mathematically the point of T is even simpler. Consider a continuous real-valued function defined on the nonnegative orthant of R^2, $f(x_1, x_2)$. Suppose one of the variables (say x_1) is fixed as a, and ϕ is defined by

$$\phi(x_2) = f(a, x_2)$$

The value of ϕ can be plotted against x_2. Then the marginal-revenue-product-of-capital curve ($MRPK$) in Figure 1 of T is a curve, where x_1 and x_2 are identified as labor and capital, respectively, and f is identified as the marginal revenue product of capital. Obviously the value of f would change when the value of x_1 changes from a to some other value, *even if the value of x_2 remains the same.* This is the "shift" of the ϕ curve, and the shift of the G_k curve in T. Such a shift of the ϕ curve will *not violate the continuity of the f function*, contrary to P-C's objection to T.

The "Averch-Johnson effect" says in essence that a firm tends to increase its investment. In order to conclude that the value of $f(x_1, x_2)$ increases, it is not sufficient to say that f is a decreasing function in x_2 for each fixed value of x_1 (see Figure 1 of T, p. 258). In T, the proof of the A-J effect is spelled out in detail in Section II, pp. 259–260. It is a usual comparative statics procedure which is familiar to economists. Somehow P-C fail to appreciate the significance of this proof as compared to the original A-J argument used to show the A-J effect. (See, for example, P-C's Figure 1, p. 212.)

II. Proof of the A-J Effect

P-C claim that the A-J effect in both T and A-J could be proved only if it is assumed to start with that the Lagrangian multiplier λ^* is strictly between zero and one. Their contention is based on the assertion that the continuity of $\lambda(s)$ at $s=0$ (where s is the upper bound on returns to capital) "... could not be proved ..." without at least one of the above assumptions. Rather than proving this impossibility, P-C claim that it is possible to prove that if the A-J effect holds or if $0<\lambda^*<1$, then the A-J effect holds. The difference should be obvious.

When $0<\lambda^*<1$, the proof of A-J effect is spelled out in Section II of T which P-C apparently ignore. In the following, we shall show that it is not necessary to assume that $0<\lambda^*<1$ nor the conclusion, in order to prove the A-J assertion.

We think that a part of their error may be caused by the notation. We shall denote the return on capital by ρ, so that:

$$\rho \equiv \frac{\pi + rK}{K}$$

where $\pi = \pi(L, K) \equiv pF(L, K) - wL - rK$, and denotes profit in T. We use superscript o to denote the unconstrained profit maximum and the asterisk to denote the constrained profit maximum. For the rest, we use the notation of T.

* Associate professor of economics, University of Kansas, and professor of economics, Purdue University, respectively. We are grateful to Professor Paul W. McAvoy for pointing out an error in an earlier draft.

Assume that

$$\pi(L, K)\big|_{L=K=0} = 0$$

and that $s>0$. Assume, following T and P-C, that $L^0>0$, $K^0>0$, $L^*>0$ and $K^*>0$.[1] Then we obtain that π^0, the unconstrained profit maximum, and π^*, the constrained profit maximum, are positive.[2] Following T, denote the marginal revenue product of labor ($MRPL$) and the marginal revenue product of capital ($MRPK$) by G_L and G_K, respectively. Write $G_L^* \equiv G_L(L^*, K^*)$, $G_K^* \equiv G_K(L^*, K^*)$, $G_L^0 \equiv G_L(L^0, K^0)$ and $G_K^0 \equiv G_K(L^0, K^0)$. Then the unconstrained case is characterized, in view of the concavity[3] of π, etc., by

$$\text{(1a)} \qquad G_L^0 = w$$

$$\text{(1b)} \qquad G_K^0 = r$$

The regulatory constraint is $\rho \leq s$. Hence:

$$\rho^* \leqq s \quad \text{where} \quad \rho^* \equiv \rho(L^*, K^*);$$

i.e.,

$$\pi^* + rK^* = \rho^* K^* \leqq sK^*,$$

and so

$$\pi^* \leqq (s-r)K^*$$

Since $\pi^*>0$ and $K^*>0$ we have:

$$\text{(2)} \qquad r < s$$

For the constrained case we have at (L^*, K^*):

$$\text{(3a)} \qquad (1-\lambda^*)G_L^* = (1-\lambda^*)w$$

$$\text{(3b)} \qquad (1-\lambda^*)G_K^* = r - \lambda^* s$$

$$\text{(3c)} \qquad \lambda^* \geqq 0 \quad \text{and}$$

$$0 = \lambda^*(sK^* + wL^* - G^*) \qquad (=\lambda^*(s-\rho^*))$$

[1] Needless to say, (L^0, K^0) denotes a solution for the unconstrained case and the (L^*, K^*) denotes a solution for the constrained case.

[2] Otherwise (L^0, K^0) and (L^*, K^*) are not solutions of the respective maximization problems, except for the trivial case. Needless to say, $L=K=0$ means "do nothing."

[3] For some discussion in which the usual first-order condition gives a necessary and sufficient characterization of optimum, see T.

Assuming that G is concave, we have:

$$\text{(4a)} \qquad G^0 - G^* \leqq G_L^* \cdot (L^0 - L^*) + G_K^* \cdot (K^0 - K^*)$$

$$\text{(4b)} \qquad G^* - G^0 \leqq G_L^0 \cdot (L^* - L^0) + G_K^0 \cdot (K^* - K^0)$$

$$\text{(4c)} \qquad G(0,0) - G^* \leqq G_L^* \cdot (0 - L^*) + G_K^* \cdot (0 - K^*)$$

where $G(0,0) = G(K, L)\big|_{K=0, L=0}$

Now $\lambda^* \neq 1$, since $\lambda^*=1$ implies, by (3b), that $r=s$ contradicting (2). Hence $(1-\lambda^*) \neq 0$, and by (3a) we have:

$$\text{(5a)} \qquad G_L^* = w$$

By (3b) we have

$$\text{(5b)} \qquad G_K^* - r = \lambda^* G_K^* - \lambda^* s$$

By (1) and (5a) we write (4a) and (4b) as:

$$\text{(6a)} \qquad G^0 - G^* \leqq w \cdot (L^0 - L^*) + G_K^* \cdot (K^0 - K^*)$$

$$\text{(6b)} \qquad G^* - G^0 \leqq w \cdot (L^* - L^0) + r \cdot (K^* - K^0)$$

By our assumption that $\pi(0,0)=0$, we have $G(0,0)=0$. We write (4c) as:

$$G^* \geqq G_L^* L^* + G_K^* K^*$$

Thus, by (5a), we have:

$$\text{(6c)} \qquad G^* \geqq wL^* + G_K^* K^*$$

Adding (6a) and (6b) we get:

$$\text{(7)} \qquad 0 \leqq (G_K^* - r)(K^0 - K^*)$$

By (7), $K^* \geqq K^0$ if, and only if, $G_K^* - r \leqq 0$.

We further, by (3c), have:

$$\lambda^* sK^* = \lambda^* G^* - \lambda^* wL^*$$

Thus, by (5b), $(G_K^* - r)K^* = \lambda^* G_K^* K^* + \lambda^* wL^* - \lambda^* G^*$. But, by (6c) and $\lambda^* \leqq 0$, $\lambda^* G_K^* K^* + \lambda^* wL^* - \lambda^* G^* \leqq 0$. Thus, since $K^*>0$, $G_K^* - r \leqq 0$. Hence, $K^* \geqq K^0$, which

is the "weak" A-J effect. If we assume $K^* \neq K^0$ then $K^* > K^0$ which is the "strong" A–J effect.

III. Relation to $0<\lambda^*<1$

The above argument shows that the A-J assertion could be proved *without assuming that* $0<\lambda^*<1$. To show that $\lambda^*>0$ we need only to assume that $K^0 \neq K^*$ and that (1) has a unique solution.[4] Indeed, if $\lambda^*=0$ then (3) becomes $G_L^*=w$ and $G_K^*=s$. It is true that $\lambda^*<1$ if, and only if, $G_K^* \leq r$, i.e., if, and only if, $K^* \geq K^0$ by (7). This is true since, by (2) and (3b):

$$(1-\lambda^*)G_K^* < (1-\lambda^*)s$$

Recalling that $\lambda^* \neq 1$, we have: $(1-\lambda^*)(G_K^*-s)<0$. Also, by (5b)

$$G_K^* - r = \lambda^*(G_K^* - s)$$

Thus P-C are "almost" correct in concluding that $0<\lambda^*<1$ if, and only if, $K^*>K^0$. But they are wrong in stating that the A-J effect is impossible to prove without assuming A-J effect or $0<\lambda^*<1$.[5] In other words, one can prove that $\lambda^*>0$ without assuming the A-J effect and one can prove the A-J effect without assuming anything about λ^*.

[4] From an elementary knowledge of the non-linear programming theory with inequality constraints λ^* as defined in T is restricted as $\lambda^* \geqq 0$ (see condition f in T (p. 256)). This remark is probably necessary in view of equation (15) of P-C.

[5] See P-C's argument leading to their equation (17), p. 212. Incidentally, their argument (apparently using their Figure 1) falls into the error of confusing the shift of the curve and the movement along the curve, which T pointed out.

REFERENCES

H. Averch and L. O. Johnson, "Behavior of the Firm Under Regulatory Constraint," *Amer. Econ. Rev.*, Dec. 1962, *52*, 1052–69.

I. Pressman and A. Carol, "Behavior of the Firm Under Regulatory Constraint: Note," *Amer. Econ. Rev.*, Mar. 1971, *61*, 210–12.

A. Takayama, "Behavior of the Firm Under Regulatory Constraint," *Amer. Econ. Rev.*, June 1969, *59*, 255–60.

Behavior of the Firm Under Regulatory Constraint: Reply

By Israel Pressman and Arthur Carol*

We completely agree that it would be useful to clarify the papers by Harvey Averch and Leland Johnson, and by Akira Takayama. Rather than clarify, however, Mohamed El-Hodiri and Takayama (E-T) have tried to reinforce their arguments to prove an unproveable result and have resorted to the same assumptions and methods of the paper by Takayama.

It is apparent that E-T have confused a function of one variable with a function of two variables. Clearly if the value of a changes then it is not the value of f that changes, but the entire function $f(a, x_2)$ becomes a new function. Furthermore, E-T continue to misinterpret the "A-J effect" as "in essence that a firm tends to increase its investment." This is not the effect that A-J attempt to prove. A-J's statement of the overcapitalization effect is that "The firm will substitute capital for the other factor of production and operate at an output where cost is not minimized" (p. 1053). In other words, the optimum capital-to-labor ratio for the regulated firm will exceed the optimum capital-to-labor ratio for the unregulated firm. The result that $K^* > K^0$ is a foregone conclusion and is implied by the assumptions of the model and does not have to be proved.

It has been shown by L. Courville that a proof of overcapitalization requires the additional assumption of strictly convex isoquants. That this is so can be seen from the diagrams of E. E. Zajac, p. 120, that require the isoquants to intersect the constraint curve in two points. This is possible only if the isoquants are strictly convex. Thus, Courville has strengthened our contention "that the very assumptions used to prove the A-J Effect . . . require an assumption that the A-J Effect exists in the first place" (p. 210).[1]

With regard to T's comparative statics procedure, p. 259–260, it should be pointed out that comparative statics should be applied very carefully and only in the vicinity of the optimum. Incorrectly applied it leads to unsound conclusions. For example, equation (23) of T, p. 259, implies that

$$dk^*/k^* = ds/(r - r_0)$$

or that the change in capital is independent of the $MRPK$, a very unsound economic implication.

Finally, it should be pointed out that strict concavity of any function doesn't imply that the second derivative is negative everywhere. For example the function $f(X) = -X^4$ is strictly concave, but $f''(0) = 0$.

REFERENCES

H. Averch and L. L. Johnson, "Behavior of the Firm Under Regulatory Constraint," *Amer. Econ. Rev.*, Dec. 1962, *52*, 1052–69.

L. Courville, uncompleted Ph.D. dissertation at Carnegie-Mellon University on Rate-of-Return Regulation.

M. El-Hodiri and A. Takayama, "Behavior of the Firm Under Regulatory Constraint: Clarifications," *Amer. Econ. Rev.*, Mar. 1973, *63*, 235–37.

I. Pressman and A. Carol, "Behavior of the Firm Under Regulatory Constraint: Note," *Amer. Econ. Rev.*, Mar. 1971, *61*, 210–12.

A. Takayama, "Behavior of the Firm Under Regulatory Constraint," *Amer. Econ. Rev.*, June 1969, *59*, 255–60.

E. E. Zajac, "A Geometric Treatment of Averch-Johnson's Behavior of The Firm Model," *Amer. Econ. Rev.*, Mar. 1970, *60*, 117–25.

* Assistant professor of statistics and associate professor of economics at Baruch College, City University of New York, respectively. Pressman is also a member of the Management Sciences Division of the American Telephone and Telegraph Company.

[1] These comments are solely the authors' and are based on interpretations of discussions with Courville. They should not be represented as his opinion.

NOTES

A new organization, the History of Economics Society, has been organized for the purpose of stimulating interest and research in the history of economics, encouraging communication and interaction among scholars working in the field, and exposing members of the profession, particularly younger ones, to many facets of the broad scientific-literary-philosophic tradition of our discipline.

The members of the Executive Committee are William Allen, Robert Eagly, Craufurd Goodwin, William Grampp, William Jaffé, Warren Samuels, Joseph Spengler, George Stigler, and Vincent Tarascio. Anyone interested in participating as a member of the Society should communicate directly with Vincent J. Tarascio, Chairman, Executive Committee, History of Economics Society, Economics Department, University of North Carolina, Chapel Hill, North Carolina 27514.

Invest-in-America announces an essay contest open to members of economic faculties. The essay should identify current problems of the American economy, and the essayist should tell what he foresees as the major economic problems of the next ten years. Essays should not exceed three thousand words, double spaced typed, and should be submitted by Sept. 1, 1973. Send manuscripts to Invest-in-America National Council, Inc., Suite 906-907, 117 South 17th Street, Philadelphia, Pennsylvania 19103. Prizes of $250.00, $150.00, and $100.00 will be announced May 1, 1974.

New Journals

The *Industrial Organization Review* will begin publication with the spring issue in 1973. The *Review* will be published initially three times a year. It will be devoted to technical and empirical papers in the area of applied microeconomic theory, as well as studies of industry structure, conduct, and performance. It will also carry occasional papers which deal with policy questions of substantial significance. Manuscripts should be submitted to Managing Editor, *Industrial Organization Review*, Building 276, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061.

The *Journal of European Economic History* is an international scientific journal addressed to the community of scholars in economic history. It is published in the English language and appears three times a year. The editor is Luigi de Rosa, formerly professor of economic history at the University of Bari, and now at Naples. The American members of the editorial board are Shepard B. Clough and Ira Glazier with offices at Temple University. Contributions and suggestions for articles and book reviews are invited. The *Journal* is published by the Banco di Roma, Via del Corso, Rome. No charge is made for subscriptions to institutions and professional economic historians.

The *Home Economic Research Journal* is a new publication of the American Home Economics Association, a professional organization dedicated to the improvement of the quality and standards of individual and family life. The *Journal* provides a medium for reporting and recording the significant research that is being conducted today in home economics. All material will be refereed and must meet high standards, both in the research and in its presentation. Manuscripts may be submitted to the editor, Geitel Winakor, MacKay Hall, Iowa State University, Ames, Iowa 50010. A guide for authors may be obtained from the editor or from the American Home Economics Association, 2010 Massachusetts Ave., N.W., Washington, D.C. 20036.

A three-week Short Course for college teachers of economics will be held at Northern Illinois University from August 6–24, 1973. The workshop is supported by a grant from the National Science Foundation. The program will acquaint teachers of undergraduates with a wide range of computer-assisted instruction (CAI) techniques in economics. Thirty participants will be selected from junior and community colleges, four-year undergraduate institutions, and universities. For application forms and further information, write Professor John C. Soper, Department of Economics, Northern Illinois University, DeKalb, Illinois 60115.

A National Symposium on Ultimate Disposal of Wastewaters and Their Residuals will be held in Durham, North Carolina on April 26 and 27, 1973. Sponsored by the Research Triangle Universities and several national organizations, the program will include sessions on land disposal, marine disposal, sludge handling, design practice, recovery, and recycling. For further information, please contact F. E. McJunkin, Associate Director, Water Resources Research Institute of the University of North Carolina, North Carolina State University, 124 Riddick Building, Raleigh, North Carolina 27607.

The Committee on the Undergraduate Program in Mathematics (Mathematical Association of America) has recently published a report entitled *Introductory Statistics Without Calculus*. Intended as a guide for departments which offer a basic service course in statistics for students who have not studied calculus, this document contains recommended objectives for such a course and a discussion of their implications. A detailed outline is given for a conventional course in introductory statistics, while some alternate approaches (decision-theoretic, nonparametric, problem-oriented) are also discussed. The report considers the use of data sets, computers, experiments, and various teaching aids. It contains a selected bibliography of almost 100 titles. Copies of the report are available without charge from the CUPM Central Office, P.O. Box 1024, Berkeley, California 94701.

The University of Denver Graduate School of International Studies announces a conference on Comparative Public Policy: Process, Output, and Tradeoffs to be held November 21–23, 1973. Papers are being solicited and younger scholars are encouraged to make submissions. Papers must involve either cross-national or cross-issue comparisons. Preference will go to papers dealing with allocational decision making or analysis of output regarding defense and/or welfare. Authors whose papers are accepted for delivery will receive an honorarium of $300 and consideration for publication. For further information, write to William Loehr, Assistant Dean, Graduate School of International Studies, University of Denver, Denver, Colorado 80210.

The American Academy of Political and Social Science recently assembled a group of leading American social scientists who conferred for two days on the subject, "Integration of the Social Sciences through Policy Analysis." Three principal papers and nine written critiques were prepared and published in a monograph. Anyone wishing to have a copy of this monograph may obtain one free by writing to this Academy at 3937 Chestnut Street, Philadelphia, Pennsylvania 19104.

A complete set of *The American Economic Review* through 1950 is available as a donation to a small college or foreign institution. For information, write Professor Robert R. Dince, Department of Banking and Finance, Varsity, Building, University of Georgia, College of Business Administration, Athens, Georgia 30601.

The fourth Congress of the International Economic Association will be held in Budapest from 19 to 25 August 1974 on the general theme, *Economic Integration: Regional and Sectoral*. The program committee consists of Fritz Machlup, Chairman, Bela Balassa (*US*), Raymond Barre (France), O. T. Bogomolov (*USSR*), Bela Czikos-Nagy (Hungary), Robert Gardiner (Africa), Herbert Giersch (West Germany), Harry Johnson (*UK*), Ryitaro Komiya (Japan), Erik Lundberg (Sweden), Jozef Pajestka (Poland), and Raul Prebisch (Latin America). The program will include plenary sessions as well as meetings of ten separate working groups dealing with various aspects of the main theme. Details will be announced within the next few months. The address of the IEA Secretariat is 54 Boulevard Raspail, 75270 Paris Cedex 06 (Secretary General—Professor Luc Fauvel). Professor Mihaly Simai, Hungarian Economic Association, Kossuth Lajos Ter. 4, Budapest V, is in charge of local arrangements.

Fulbright-Hays Program for Senior American and Foreign Scholars in Economics and Business Administration

Applications will be accepted this spring for more than 550 lecturing and advanced research awards during 1974–75 in over 75 countries under the senior Fulbright-Hays program, the Committee on International Exchange of Persons announced recently. Specialists in economics who are U.S. citizens and have a doctorate or college teaching experience are invited to indicate their interest in an award by completing a simple registration form, available on request from Senior Fulbright-Hays Program, 2101 Constitution Avenue, Washington, D.C. 20418. Registrants will receive a detailed announcement of the 1974–75 program in May. The deadline for applying for research awards is July 1, 1973 and it is also the suggested date for filing for lectureships.

Applications from senior foreign scholars for temporary appointments at American colleges or universities are transmitted to the Committee each year by Fulbright-Hays agencies abroad. The scholars are eligible for a Fulbright-Hays travel grant upon receiving a lecturing or research appointment. An annual list of such scholars is issued in March. Also available is a list of 26 senior Fulbright-Hays foreign scholars in economics who are in the United States this academic year. A number of them would be pleased to accept invitations to give some lectures or to participate in special conferences under the sponsorship of academic institutions and educational organizations.

Deaths

Leland Louis Howell, Reno, Nevada, Nov. 20, 1972.

Retirements

George H. Hand, professor of economics, Southern Illinois University, Sept. 1, 1971.

Taylor D. MacLafferty, associate professor of business administration, Emory University, Sept. 1, 1972.

Vernon Morrison, professor of economics, Southern Illinois University, June, 1972.

Visiting Foreign Scholars

Do Yung Chung, Sung Kyun Kwan University: visiting professor of economics, Thunderbird Graduate School, fall 1972.

Nicos E. Devletoglou, London School of Economics and Political Science, University of London: Texas A&M University, spring 1973.

Promotions

Robert F. Allen: associate professor of economics, University of Nebraska, 1972–73.

Evan E. Anderson: associate professor, Graduate School of Business Administration, Tulane University, July 1, 1972.

A. Paul Ballantyne: professor, College of Letters, Arts, and Sciences, University of Colorado, June 1972.

Douglas Bohi: associate professor of economics, Southern Illinois University.

E. Gerald Corrigan: senior economist and assistant secretary, research and statistics function, Federal Reserve Bank of New York.

William P. Culbertson: assistant professor, department of economics, University of Virginia, Feb. 1. 1972.

Lawrence A. Daellenbach: associate professor, University of Wisconsin, La Crosse, Aug. 28, 1972.

Don J. DeVoretz: associate professor, department of economics and commerce, Simon Fraser University, Sept. 1, 1971.

James Fenner: associate professor of economics, University of Bridgeport, Sept. 1972.

Richard Fryman: associate professor of economics, Southern Illinois University, July 1971.

William R. Graham: associate professor of economics and statistics, California State University, Los Angeles, Sept. 1972.

Margaret L. Greene: manager, foreign department, Federal Reserve Bank of New York.

Andrew C. Gross: associate professor, Cleveland State University, Sept. 1, 1971.

Harish C. Gupta: associate professor of economics, University of Nebraska, 1972–73.

Michael J. Hamburger: senior economist, research and statistics function, Federal Researve Bank of New York.

Judd Hammack: associate professor of economics and statistics, California State University, Los Angeles, Sept. 1972.

William B. Hartley: professor of economics, Iowa Wesleyan College, Sept. 1, 1972.

M. F. Hassan: professor of economics, Illinois State University.

F. Gregory Hayden: associate professor of economics, University of Nebraska, 1972–73.

Paul L. Helsing: professor, department of economics, Eastern Washington State College, May 1972.

Paul Heyne: professor, department of economics, Southern Methodist University, May 1972.

M. Bruce Johnson: professor of economics, University of California, Santa Barbara, July 1972.

Dennis R. Maki: associate professor, department of economics and commerce, Simon Fraser University, Sept. 1, 1971.

Lloyd J. Mercer: associate professor of economics, University of California, Santa Barbara, July 1972.

Turley Mings: professor, economics department, California State University, San Jose, Sept. 1, 1972.

Paul T. Mu: associate professor of economics, California State University, Los Angeles, Sept. 1972.

Edgar O. Olsen: associate professor, department of economics, University of Virginia, Sept. 1, 1972.

Virginia L. Owen: associate professor of economics, Illinois State University.

Ramachandra Ramanathan: associate professor of economics, University of California, San Diego, July 1, 1972.

Edward J. Regan: chief, financial statistics division, Federal Reserve Bank of New York.

Nagesh S. Revankar: associate professor of economics, State University of New York at Buffalo.

Milton Russell: professor of economics, Southern Illinois University, June 1971.

William B. Simpson: professor of economics, California State University, Los Angeles, Sept. 1972.

Russell A. Snyder: professor of economics, Eastern Washington State College, Sept. 1972.

Kevin C. Sontheimer: associate professor of economics, State University of New York at Buffalo.

Gerald F. Sorrensen: associate professor of economics, California State University, Los Angeles, Sept. 1972.

Gary Stern: chief, business conditions division, Federal Reserve Bank of New York.

Robert C. Stuart: associate professor of economics, Douglass College, Rutgers University, July 1, 1972.

Sam Sydneysmith: associate professor, department of economics and commerce, Simon Fraser University, Sept. 1, 1972.

Arlon R. Tussing: professor, University of Alaska, July 1972.

Harold L. Votey, Jr.: associate professor of economics, University of California, Santa Barbara, July 1972.

William C. Wedley: assistant professor, department of economics and commerce, Simon Fraser University, August 25, 1971.

Administrative Appointments

Alec P. Alexander: dean, College of Letters and Science, University of California, Santa Barbara, July 1971.

A. Paul Ballantyne: dean, College of Letters, Arts, and Sciences, University of Colorado, Apr. 1972.

Parzival Copes: chairman, department of economics and commerce, Simon Fraser University, Sept. 1, 1972.

John Drake, University of Washington: chairman, department of world business, Thunderbird Graduate School, Sept. 1972.

John S. Dydo: director, administrative studies program, Southampton College of Long Island University, July 25, 1972.

Edna E. Ehrlich: manager, international research department, Federal Reserve Bank of New York.

Marshall I. Goldman: chairman, department of economics, Wellesley College, Sept. 1971.

William B. Hartley: chairman, department of economics and business administration, Iowa Wesleyan College, Sept. 1, 1972.

John B. Henderson: chief, economics division, Congressional Research Service, Library of Congress, Sept. 1972.

Frank I. Jewett: associate dean for instructional resources, California State University, Humboldt, July 1, 1972.

Stanley R. Johnson: chairman, department of economics, University of Missouri, Columbia, July 1, 1972.

Carl R. Jones: associate chairman (academic planning and instruction), department of operations research and administrative sciences, Naval Postgraduate School, July 1971.

James V. Koch: chairman, department of economics, Illinois State University.

Stephen J. Kohn, Akron National Bank & Trust Company: superintendent of banks, State of Ohio, Nov. 1, 1972.

David H. Kopf: chief, money and finance division, Federal Reserve Bank of New York.

Cletus H. Kruyer, Jr. vice-president of college affairs, High Point College.

Robert G. Layer: chairman, department of economics Southern Illinois University, Feb. 1, 1972.

Lloyd J. Mercer: vice-chairman, department of economics, University of California, Santa Barbara, July 1972.

Turley Mings: director, Center for Economics Education, California State University, San Jose, Sept. 1, 1972.

Gordon A. Philpot: chairman, department of economics, Whitman College, June 1, 1972.

Alfred E. Pierce: acting head, department of economics and business, Lafayette College, Aug. 1, 1972.

John E. Rapp, MacMurray College: professor of economics and chairman, University of Dayton, Aug. 1972.

George Rosen: head, department of economics, University of Illinois, Chicago Circle, Sept. 1972.

John B. Ross: vice-president, Bank of America, San Francisco, Oct. 18, 1972.

William R. Russell: chairman, department of economics, Southern Methodist University, June 1972.

George G. Sause: provost, Lafayette College, Aug. 1, 1972.

Richard T. Selden: chairman, department of economics, University of Virginia, Sept. 1, 1972.

Malcolm F. Severance, acting chairman, department of economics and business administration, University of Vermont, July 1, 1972.

Jerome B. Siebert: associate director, Agricultural Extension, University of California at Berkeley, Sept. 1, 1972.

Russell A. Snyder: chairman, department of economics, Eastern Washington State College, Sept. 1972.

Douglas P. Sweetland: department chairman, department of economics, University of Wisconsin, La Crosse, Aug. 28, 1972.

Harold L. Votey, Jr.: acting director, Center for the Study of Developing Nations and the Community and Organizations Research Institute, University of California, Santa Barbara, July 1971.

Richard L. Wallace: director of Policy Research Center and assistant dean for research, College of Administration and Public Affairs, University of Missouri, Columbia.

W. Donald Wood, director, Industrial Relations Center, Queen's University: acting chairman, Ontario Economic Council.

New Appointments

Roy D. Adams: assistant professor of economics, Iowa State University.

Rahim Amin, University of Cincinnati: assistant professor of economics, Anderson College, Sept. 1972.

Geoffrey T. Andron: visiting assistant professor of economics, Iowa State University, 1972–73.

Richard K. Armey, Austin College: associate professor of economics, North Texas State University, Sept. 1, 1972.

Richard W. Ault, University of Virginia: lecturer, Texas A&M University.

Raveendra N. Batra, University of Western Ontario: associate professor, department of economics, Southern Methodist University, Sept. 1972.

Edward I. Beck: instructor, department of economics and business, Lafayette College, Sept. 1. 1972.

R. A. Douglas Beck, Iowa State University: assistant professor, department of economics and commerce, Simon Fraser University, Sept. 28, 1972.

Sohrab Behdad, Michigan State University: lecturer, Syracuse University, 1972–73.

David A. Belsley: senior research associate, National Bureau of Economic Research, Computer Research Center.

Robert L. Bish, University of Washington: associate professor, department of economics and Center for Urban Affairs, University of Southern California, Sept. 1972.

Tapan K. Biswas: lecturer, department of economics, University of California, Santa Barbara, July 1972.

Charles Blackorby, London School of Economics: associate professor of economics, Southern Illinois University, Sept. 1972.

Michael K. Block, University of Santa Clara: assistant professor of economics, Naval Postgraduate School, Sept. 1972.

Gregory S. Boussios: assistant professor of economics, Southampton College of Long Island University, Sept. 1, 1972.

D. L. Brito, University of Wisconsin: associate professor, The Ohio State University.

Robert Brownlee: instructor, Syracuse University, Sept. 1972–73.

David C Bunting: visiting assistant professor, department of economics, Eastern Washington State College, Sept. 1972.

Stephen Buser, Boston College: assistant professor of economics, Southern Illinois University, Sept. 1972.

Tansu U. Ciller, University of Connecticut: assistant professor of economics, Franklin and Marshall College, Sept. 1972.

Charles E. Conrod, Northwestern University: visiting instructor, Simon Fraser University, Sept. 1, 1972.

John S. Cook, Purdue University: assistant professor, department of economics, University of South Carolina.

L. Jorn Dakin, University of Kansas: assistant professor of economics and business law, University of North Florida, Sept. 1972.

Martin Davis: assistant professor, economics department, California State University, San Jose, Sept. 1, 1972.

Robert T. Deacon: assistant professor of economics, University of California, Santa Barbara, July 1972.

Merrill E. Douglass: assistant professor of business administration, Emory University, Sept. 1, 1972.

Douglas Dowd: lecturer, economics department, California State University, San Jose, Sept. 1, 1972.

John S. Dydo, C.R. Dean Economics, Inc.: professor of economics, Southampton College of Long Island University, July 1, 1972.

Randyl D. Elkin: visiting assistant professor of economics, Iowa State University, 1972–73.

Howard Elliott, Princeton University: research associate, Center for Research on Economic Development, University of Michigan.

James Elliott, Yale University: research associate,

Center for Research on Economic Development, University of Michigan.

Thomas W. Epps, U.S. Army: assistant professor, department of economics, University of Virginia, Sept. 1, 1972.

Teviah L. Estrin: instructor, department of economics and commerce, Simon Fraser University, Sept. 1, 1972.

Bert Evans: associate professor of economics, University of Nebraska, 1972–73.

Roger N. Folsom, Claremont Graduate School: assistant professor of economics, Naval Postgraduate School, Sept. 1972.

Ralph R. Frasca, Indiana University: assistant professor of economics, University of Dayton, Aug. 1972.

Stuart A. Gallaher: instructor, department of economics and business, Lafayette College, Sept. 1, 1972.

Gillian G. Garcia: assistant professor, economics department, California State University, San Jose, Sept. 1, 1972.

Marshall Geer, University of Colorado: associate professor of economics, Thunderbird Graduate School, Sept. 1972.

Charles G. Geiss: assistant professor of economics, University of Missouri, Columbia.

Betty Gibson: lecturer, economics department, California State University, San Jose, Sept. 1, 1972.

Errol Glustoff, University of Tennessee: visiting associate professor, department of economics, University of California, Santa Barbara, 1972–73.

Robert S. Golden: assistant professor of economics and statistics, California State University, Los Angeles, Sept. 1972.

Loretta L. Good: assistant professor, University of Wisconsin, La Crosse, Aug. 1972.

Hans-Werner Gottinger: lecturer, department of economics, University of California, Santa Barbara, Jan. 1973.

Levern Graves, California State University, Fullerton: visiting professor, California State University, Los Angeles, summer 1972.

Ronald E. Grieson, Massachusetts Institute of Technology: assistant professor, Queens College, City University of New York, 1972–73.

Reuben Gronau: research associate, National Bureau of Economic Research, June 1972.

Arthur Guthrie, Carleton University: assistant professor, department of economics and commerce, Simon Fraser University, Jan. 1, 1973.

Josef Hadar, Case Western Reserve University: professor, department of economics, Southern Methodist University, Sept. 1972.

Gus W. Haggstrom, University of California, Berkeley: staff member, economics department, The Rand Corporation, Aug. 1972.

Bennett Harrison, University of Maryland: associate professor of economics and urban studies, Massachusetts Institute of Technology, Mar. 1973.

Jacob Hornik: assistant professor, department of marketing, University of Illinois at Chicago Circle.

Ernest L. Inwood, California State University, San Jose: professor, economics department, Eastern College, Sept. 1, 1972.

J. Mark Jackson, Auburn University: visiting assistant professor, Texas A&M University.

John A. James, Massachusetts Institute of Technology: acting assistant professor, department of economics, University of Virginia, Feb. 1, 1973.

Karen Johnson, Massachusetts Institute of Technology: instructor, Wellesley College, Sept. 1971.

William A. Johnson, Rand Corporation: staff economist, Council of Economic Advisers.

Suhas Ketkar, Vanderbilt University: lecturer, University of Sierra Leone and research associate, Center for Research on Economic Development, University of Michigan.

Larry Kiser: assistant professor, department of economics, Eastern Washington State College, Sept. 1972.

Odin K. Knudsen: lecturer, economics department, California State University, San Jose, Sept. 1, 1972.

James V. Koch, Northern Illinois University: visiting associate professor, California State University, Los Angeles, summer 1972.

Robert Konwea, California State University, San Jose: visiting assistant professor, California State University, Los Angeles, summer 1972.

John D. Kyle: visiting associate professor, department of economics and commerce, Simon Fraser University, Sept. 1, 1972.

William K. Kyle: assistant professor of administration, Sangamon State University, fall 1972.

Helen Ladd, Harvard University: instructor, Wellesley College, Sept. 1971.

Harvey E. Lapan: assistant professor of economics, Iowa State University.

Norman B. Lefton: consultant economist in medical affairs, Regional Medical Program of Hawaii, Sept. 1972.

Roger C. Leonhard: instructor, department of economics, University of Virginia, Sept. 1, 1972.

Dianne R. Levene: assistant professor of economics, California State University, Los Angeles, Sept. 1972.

Wilbur C. Lewellen: research associate, National Bureau of Economic Research, July 1972.

Lee A. Lillard: research associate, National Bureau of Economic Research, June 1972.

Thomas J. Lombardo, Texas Tech University: assistant director, Center for Economic Education, North Texas State University, July 1, 1972.

Edwin J. Longfellow: instructor, department of economics, Iowa State University.

Richard E. Low, Rutgers University: visiting associate professor, California State University, Los Angeles, summer 1972.

Dennis A. MacDonneil, University of California, Riverside: international training officer, Bank of America, July 1972.

Peter J. McCabe: acting assistant professor of economics, University of California, Santa Barbara, July 1971.

John M. Marshall, Vanderbilt University: assistant professor of economics, University of California, Santa Barbara, July 1972.

James C. Miller III, U.S. Department of Transporta-

tion: associate professor, Texas A&M University.

Kenneth E. Moberg: assistant professor of business administration, Emory University, Sept. 1, 1972.

Terry Monson, University of Minnesota: research associate, Center for Research on Economic Development, University of Michigan.

Peter Moock, Columbia University: research associate, Center for Research on Economic Development, University of Michigan.

R. Allen Moran: assistant professor of economics, University of Bridgeport, Sept. 1972.

James A. Morris: professor, department of economics, University of South Carolina.

Alexander L. Morton: research associate, National Bureau of Economic Research, Aug. 1972.

Milton Moss: senior research associate, National Bureau of Economic Research, May 1972.

Dennis C. Mueller: research associate, National Bureau of Economic Research, Aug. 1972.

Paul Munyon, Harvard University: instructor, Wellesley College, Sept. 1971.

Dennis D. Olson: assistant professor of economics, Emory University, 1972–73.

Gerald W. Olson: instructor, department of economics, University of Missouri, Kansas City, Sept. 1, 1971.

Andrew Parnes: lecturer, economics department, California State University, San Jose, Sept. 1, 1972.

Newman S. Peery: instructor, department of economics and commerce, Simon Fraser University, Sept. 1, 1972.

Neal Pepper, economist, money and finance division, Federal Reserve Bank of New York.

Lawrence T. Pinfield, Stanford University: assistant professor, department of economics and commerce, Simon Fraser University, Sept. 1, 1972.

Perry W. Polk: instructor, department of economics and management, USAF Academy, June 1972.

Nikolai Pulchritudoff: assistant professor of economics and statistics, California State University, Los Angeles, Sept. 1972.

Theodore E. Quast, Jr., Indiana University, South Bend; assistant professor of economics, Marshall University, Aug. 28, 1972.

Satish Raichur, University of Pittsburgh: visiting assistant professor, Graduate School of International Studies, University of Denver.

Samuel D. Ramenofsky: assistant professor, department of economics, University of Missouri, Kansas City, Sept. 1, 1972.

Clyde G. Reed, University of Washington: assistant professor, department of economics and commerce, Simon Fraser University, Sept. 1, 1972.

Edward H. Robb: assistant professor of economics and director of the State and Regional Fiscal Studies Unit, University of Missouri, Columbia, July 1, 1972.

John D. Rossi: assistant professor, department of economics and business, Lafayette College, Sept. 1, 1972.

R. S. Rungta: visiting associate professor of marketing, Temple University, 1972–73.

Gertrude E. Schroeder: professor, department of economics, University of Virginia, Sept. 1, 1972.

Richard Schwindt: instructor, department of economics and commerce, Simon Fraser University, Sept. 1, 1972.

Ernest M. Scott: instructor, department of economics and commerce, Simon Fraser University, Sept. 1, 1972.

Gerald W. Scully, Southern Illinois University: associate professor, department of economics, Southern Methodist University, Sept. 1972.

M. Sabry El Shabrawy, University of Illinois: assistant professor of world business, Thunderbird Graduate School, Sept. 1972.

Jeremy F. Shapiro: senior research associate, National Bureau of Economic Research, Computer Research Center, July 1972.

Kenneth Shapiro, Stanford University: research associate, Center for Research on Economic Development, University of Michigan.

Stephen L. Shapiro, University of South Carolina: assistant professor of economics, University of North Florida, Aug. 1972.

James F. Smith: research associate, National Bureau of Economic Research, Oct. 1972.

John C. Soper, University of Missouri, Columbia: assistant professor of economics and coordinator, office for economic education, Northern Illinois University, Sept. 1972.

Gayle Southworth: lecturer, economics department, California State University, San Jose, Sept. 1, 1972.

Earl O. Stephens: assistant professor of economics and statistics, California State University, Los Angeles, Sept. 1972.

Wayne L. Stevenson: assistant professor, department of economics, Middlebury College, Sept. 1972.

George J. Stigler: senior research staff, National Bureau of Economic Research, July 1972.

Raymond J. Struyk, Rice University: Housing Studies Group, The Urban Institute, May 1972.

William J. Taylor: lecturer in business administration, Emory University, 1972–73.

Katsuaki Terasawa, California Institute of Technology: assistant professor of economics, Naval Postgraduate School, June 1972.

Harriet Tolpin, Boston College: instructor, Wellesley College, Jan 1973.

Edward Tower, University of Auckland: assistant professor, department of economics and commerce, Simon Fraser University, Sept. 1, 1972.

Andris Trappens, University of California, Berkeley: assistant professor of economics, Thunderbird Graduate School, Feb. 1973.

Roger Van Tassel, Clark University: visiting professor, California State University, Los Angeles, summer 1972.

Roger C. Vergin, University of Washington: professor, department of economics and commerce, Simon Fraser University, Sept. 1, 1972.

Robert Vogel, Wesleyan University: associate professor of economics, Southern Illinois University, Sept. 1972.

Dale Vorderlandwehr, Michigan State University: assistant professor of economics, Thunderbird Graduate School, Sept. 1972.

Janice Weaver, University of Illinois: lecturer, Syracuse University, 1972–73.

Rodger Weaver: assistant professor of economics, University of Nebraska, 1972–73.

Charles L. Weber: lecturer, department of economics, University of Illinois, Chicago Circle.

William E. Wehrs: assistant professor, University of Wisconsin, La Crosse, Aug. 1972.

Peter Weisel, University of Oregon: research associate, Center for Research on Economic Development, University of Michigan.

Joseph Weiss, Louis Berger, Inc.: marketing economist, Companhia Brasileira de Alimentos, Brasilia, Oct. 15, 1972.

Donald R. Winkler: assistant professor of economics, University of California, Santa Barbara, July 1972.

Louis A. Woods, East Carolina University: assistant professor of geography and economics, University of North Florida, fall 1972.

Paul Zarembka, University of California, Berkeley: associate professor of economics, State University of New York at Buffalo.

Seid M. Zekavat, Loyola University of Los Angeles: visiting associate professor, California State University, Los Angeles, summer 1972.

Leaves for Special Appointments

Albert Y. Badre, Southern Illinois University: professor of economics, American University, Beirut, 1971–72.

Abraham S. Becker, The Rand Corporation: visiting professor of economics, Soviet and East European Research Center, Hebrew University, 1972–73.

James W. Christian, Iowa State University: Federal Home Loan Bank Board, Washington, Sept. 1, 1972–Aug. 31, 1973.

Michael R. Darby, Ohio State University: visiting assistant professor, University of California, Los Angeles.

William B. Eastlake, Boise State College: Fulbright lecturer in economics, Cuttington College, Monrovia, Liberia, July 1971–July 1973.

Milton Edelman, Southern Illinois University: Fulbright research scholar, Industrial Relations Research Unit, Social Science Research Council, University of Warwick, England, Jan.–June 1973.

Julian M. Greene, Southern Methodist University: Ernst and Ernst, MCS Division, Washington, Aug. 1972.

Geoffrey B. Hainsworth, University of British Columbia: research associate, International Development Research Council, Ottawa, 1972–73; visiting fellow, Overseas Development Council, Washington, Jan.–June 1973.

Yukon Huang, University of Virginia: University of Dar es Salaam, Dar es Salaam, Tanzania, 1972–74.

John Kraft, University of Florida: Brookings Institution Economic Policy Fellow, Brookings Institution and the Price Commission, June 1972.

Fred D. Levy, Jr., Syracuse University: U.S. Treasury, July 1972.

Katharine C. Lyall, Syracuse University: senior research associate, Center for Urban Studies, Johns Hopkins University, Sept. 1972.

Richard C. Maxon, Iowa State University: senior visiting lecturer, department of agricultural economics, University of Ibadan, Ibadan, Nigeria, Sept. 1, 1972–Aug. 31, 1973.

Walter J. Mead, University of California, Santa Barbara: chief economist, The Energy Policy Project, Washington, Sept. 1, 1972–Aug. 31, 1973.

John C. Murdock, University of Missouri, Columbia: professor of economics and adviser to the graduate dean, Massey University, New Zealand, Sept. 1972.

James R. O'Connor, California State University, San Jose: Max Planck Institut, West Germany, fall 1972.

Richard L. Peterson, Southern Methodist University: Consumer Credit and Finance Section, Division of Research and Statistics, Federal Reserve Board, Aug. 1972.

R. Robert Russell, University of California, Santa Barbara: visiting associate professor of economics, University of California, San Diego, 1972–73.

Kevin C. Sontheimer, State University of New York, Buffalo: visiting associate professor of mathematical economics, Naval Postgraduate School, 1973.

Charles Stokes, University of Bridgeport: General Accounting Office, Washington.

P. A. V. B. Swamy, Ohio State University: consultant with the Board of Governors, Federal Reserve System, Washington.

Richard S. Thorn, University of Pittsburgh: visiting professor, University of Aix-Marseille, Oct. 1972; University of Paris, Feb. 1973.

Sue Van Atta, California State University, San Jose: School of Business, University of California, Berkeley, Sept. 1, 1972.

Barbara S. Zoloth, Emory University: Fellow, Brookings Institution, 1972–73.

Resignations

Kirk Blackerby, California State University, San Jose, June 30, 1972.

Arthur E. Gandolfi, Federal Reserve Bank of New York: First National City Bank.

Donald M. Lamberton, Case Western Reserve University: University of Queensland, Jan. 1973.

David R. Miller, Syracuse University: Barton-Aschman Associates, Chicago, Sept. 1972.

Ronald L. Moomaw, University of Virginia, June 30, 1972.

David V. Pritchett, Federal Reserve Bank of New York: First National City Bank.

Rachel Strauber, Federal Reserve Bank of New York.

James B. Thornblade, Syracuse University: First National Bank of Boston, Sept. 1972.

Peter G. Van der Spek, Federal Reserve Bank of New York: Overseas Business Expansion.

NOTE TO DEPARTMENTAL SECRETARIES AND EXECUTIVE OFFICERS

When sending information to the *Review* for inclusion in the Notes Section, please use the following style:

A. Please use the following categories:

1—Deaths
2—Retirements
3—Foreign Scholars (visiting the USA or Canada)
4—Promotions
5—Administrative Appointments
6—New Appointments
7—Leaves for Special appointments (NOT Sabbati
8—Resignations
9—Miscellaneous

B. Please give the name of the individual (SMITH, John W.), his present place of employment or enrollment new title (if any), and the date at which the change will occur.

C. Type each item on a separate 3x5 card and please do not send public relations releases.

D. The closing dates for each issue are as follows: *March*, November 1; *June*, February 1; *September*, Ma *December*, August 1.

This announcement supersedes and replaces a letter which was sent annually from the managing editor's office items and information should be sent to the Assistant Editor, *American Economic Review*, Box Q, Brown Univer Providence, Rhode Island 02912.

The American Economic Review

ARTICLES

JUNE 1973

THE AMERICAN ECONOMIC ASSOCIATION

Founded in 1885

Officers

President

KENNETH ARROW
Harvard University

President-Elect

WALTER HELLER
University of Minnesota

Vice-Presidents

JAMES S. DUESENBERRY
Harvard University

LAWRENCE R. KLEIN
University of Pennsylvania

Secretary-Treasurer and Editor of the proceedings

RENDIGS FELS
Vanderbilt University

Managing Editor of The American Economic Review

GEORGE H. BORTS
Brown University

Managing Editor of The Journal of Economic Literature

MARK PERLMAN
University of Pittsburgh

Executive Committee

Elected Members of the Executive Committee

ROBERT EISNER
Northwestern University

JOHN R. MEYER
Yale University

GUY HENDERSON ORCUTT
Yale University

JOSEPH A. PECHMAN
Brookings Institution

IRMA ADELMAN
University of Maryland

ROBERT HEILBRONER
New School for Social Research

Ex Officio Members

JAMES TOBIN
Yale University

JOHN KENNETH GALBRAITH
Harvard University

• Published at George Banta Co., Inc., Menasha, Wisconsin.

• THE AMERICAN ECONOMIC REVIEW, including four quarterly numbers, the *Proceedings* of the annual meetings, and *Directory* and Supplements, is published by the American Economic Association and is sent to all members five times a year, in March, May, June, September, and December.

• Membership dues of the Association are $20.00 a year, which includes a year's subscription to both the *American Economic Review* and the *Journal of Economic Literature.* A three year junior membership of $10.00 per year is available to registered students. Subscriptions by nonmembers are $30.00 a year, and only subscriptions to both publications will be accepted. Single copies of the *Review* and *Journal* are $4.00 each. Each order for copies of either publication must also include a $.50 per order service charge. Orders should be sent to the Secretary's office, Nashville, Tennessee.

• Correspondence relating to the *Papers and Proceedings,* the *Directory,* advertising, permission to quote, business matters, subscriptions, membership and changes of address may be sent to the secretary, Rendigs Fels, 1313 21st Avenue, South, Nashville, Tennessee 37212. To be effective, notice of change of address must reach the secretary by the 1st of the month previous to the month of publication. The Association's publications are mailed by second class and are not forwardable by the Post Office.

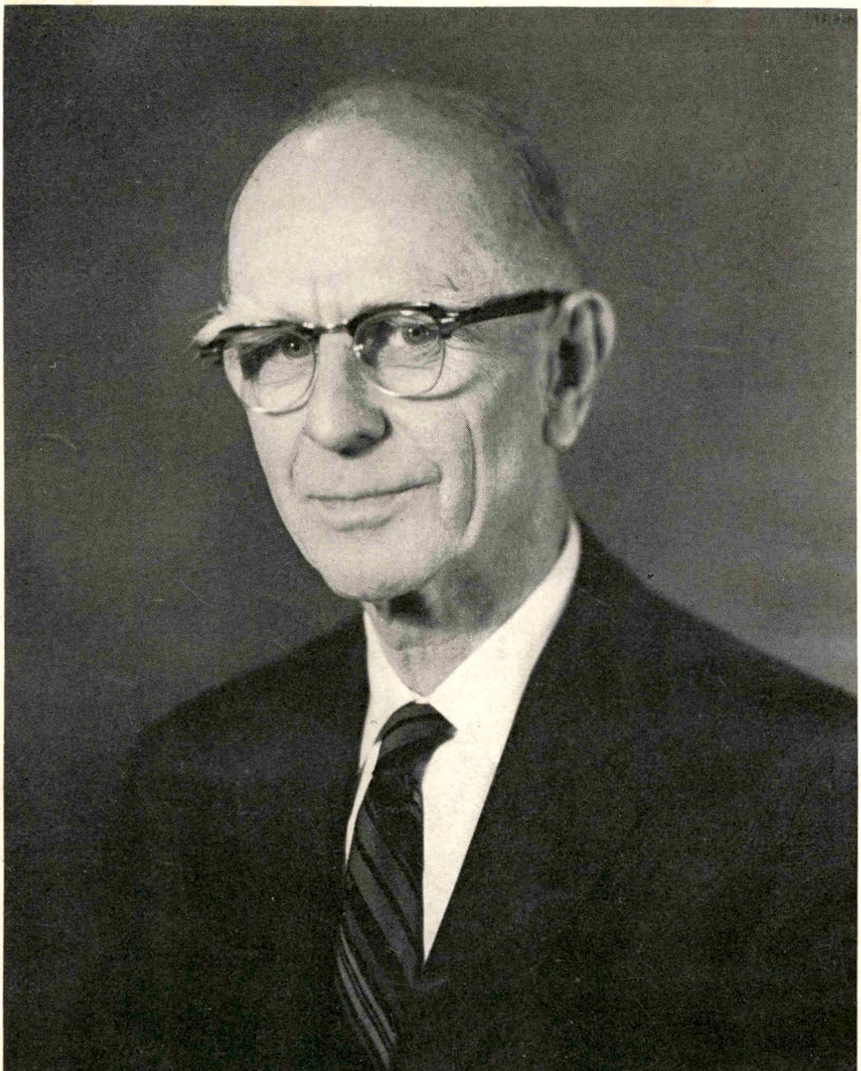
• Second-class postage paid at Nashville, Tennessee and at additional mailing offices. Printed in U.S.A.

CARL S. SHOUP

Distinguished Fellow

1972

Carl Sherman Shoup's distinguished contributions to public finance range from pure tax theory to the history of thought on taxation and practical tax reform. His reputation derives in large part from his studies of the turnover tax, tax harmonization among common market countries, and estate taxation. In addition, he was the principal architect of ingenious and detailed tax reform proposals for Japan and Venezuela. His recent treatise, *Public Finance,* is a work of major significance that brings together his contributions to economic theory and applied economics during the past forty years.



Carl S. Strong

Modern Economic Growth: Findings and Reflections

By SIMON KUZNETS*

I. Definitions

A country's economic growth may be defined as a long-term rise in capacity to supply increasingly diverse economic goods to its population, this growing capacity based on advancing technology and the institutional and ideological adjustments that it demands. All three components of the definition are important. The sustained rise in the supply of goods is the *result* of economic growth, by which it is identified. Some small countries can provide increasing income to their populations because they happen to possess a resource (minerals, location, etc.) exploitable by more developed nations, that yields a large and increasing rent. Despite intriguing analytical problems that these few fortunate countries raise, we are interested here only in the nations that derive abundance by using advanced contemporary technology—not by selling fortuitous gifts of nature to others. Advancing technology is the *permissive* source of economic growth, but it is only a potential, a necessary condition, in itself not sufficient. If technology is to be employed efficiently and widely, and, indeed, if its own progress is to be stimulated by such use, institutional and ideological adjustments must be made to effect the proper use of innovations generated by the advancing stock of human knowledge. To cite examples from modern economic growth: steam and electric power and the large-scale plants needed to exploit them are not compatible with family enterprise, illiteracy, or slavery—all of which prevailed in earlier times over much of even the developed world, and had to be replaced by more appropriate institutions and social views. Nor is modern technology compatible with the rural mode of life, the large and extended family pattern, and veneration of undisturbed nature.

The source of technological progress, the particular production sectors that it affected most, and the pace at which it and economic growth advanced, differed over centuries and among regions of the world; and so did the institutional and ideological adjustments in their interplay with the technological changes introduced into and diffused through the growing economies. The major breakthroughs in the advance of human knowledge, those that constituted dominant sources of sustained growth over long periods and spread to a substantial part of the world, may be termed epochal innovations. And the changing course of economic history can perhaps be subdivided into economic epochs, each identified by the epochal innovation with the distinctive characteristics of growth that it generated.[1] Without considering the feasibility of identifying and dating such economic epochs, we may proceed on the working assumption that modern economic growth represents such a

* Harvard University. This article is the lecture he delivered in Stockholm, Sweden, December 1971, when he received the Nobel Prize in Economic Science. The article is copyright © the Nobel Foundation 1972. It is published here with the permission of the Nobel Foundation, and is included in the volume of *Les Prix Nobel en 1971*.

[1] For a discussion of the economic epoch concept, see Kuznets (1966), pp. 1–16.

distinct epoch—growth dating back to the late eighteenth century and limited (except in significant *partial* effects) to economically developed countries. These countries, so classified because they have managed to take adequate advantage of the potential of modern technology, include most of Europe, the overseas offshoots of Western Europe, and Japan—barely one quarter of world population.[2] This paper will focus on modern economic growth, but with obviously needed attention to its worldwide impact.

Limitations of space prevent the presentation of a documented summary of the quantitative characteristics commonly observed in the growth of the presently developed countries, characteristics different from those of economic growth in earlier epochs. However, some of them are listed, because they contribute to our understanding of the distinctive problems of economic life in the world today. While the list is selective and is open to charges of omission, it includes those observed and empirically testable characteristics that lead back to some basic factors and conditions, which can only be glimpsed and conjectured, and forward to some implications that have so far eluded measurement.

II. The Six Characteristics

Six characteristics of modern economic growth have emerged in the analysis based on conventional measures of national product and its components, population, labor force, and the like. First and most obvious are the high rates of growth of per capita product and of population in the developed countries—both large multiples of the previous rates observable in these countries and of those in the rest of the world, at least until the recent decade or two.[3] Second, the rate of rise in productivity, i.e., of output per unit of all inputs, is high, even when we include among inputs other factors in addition to labor, the major productive factor—and here too the rate is a large multiple of the rate in the past.[4] Third, the rate of structural transformation of the economy is high. Major aspects of structural change include the shift away from agriculture to nonagricultural pursuits and, recently, away from industry to services; a change in the scale of productive units, and a related shift from personal enterprise to impersonal organization of economic firms, with a corresponding change in the occupational status of labor.[5]

[2] For a recent classification identifying the non-Communist developed countries, see United Nations *Yearbook*, notes to Table 5, p. 156. These classifications vary from time to time, and differ somewhat from those of other international agencies.

[3] For the non-Communist developed countries, the rates of growth per year over the period of modern economic growth, were almost 2 percent for product per capita, 1 percent for population, and 3 percent for total product. These rates—which mean roughly a multiplication over a century by five for product per capita, by three for population, and by more than fifteen for total product—were far greater than premodern rates. The latter can only be conjectured, but reasonable estimates for Western Europe over the long period from the early Middle Ages to the mid-nineteenth century suggest that the modern rate of growth is about ten times as high for product per capita (see Kuznets (1971), pp. 10–27). A similar comparison for population, either for Europe or for the area of European settlement (i.e., Europe, the Americas, and Oceania), relating to 1850–1960, as compared with 1000–1850, suggests a multiple of 4 or 5 to 1 (see Kuznets (1966), Tables 2.1 and 2.2, pp. 35 and 38). The implied acceleration in the growth rate of total product is between forty and fifty times.

[4] Using the conventional national economic accounts, we find that the rate of increase in productivity is large enough to account (in the statistical sense) for almost the entire growth of product per capita. Even with adjustments to allow for hidden costs and inputs, growth in productivity accounts for over half of the growth in product per capita (see Kuznets (1971), pp. 51–75, particularly Table 9, p. 74; and Table 11, p. 93).

[5] The rapidity of structural shifts in modern times can be easily illustrated by the changes in the distribution of the labor force between agriculture (and related industries) and the nonagricultural production sectors. In the United States, the share of labor force attached to the agricultural sector was still 53.5 percent in 1870 and declined to less than 7 percent in 1960. In an old European country like Belgium, the share of agriculture in the labor force, 51 percent in 1846, dropped to 12.5 percent in 1947 and further to 7.5 percent in 1961 (see Bairoch et al., Tables D-4 and C-4). Considering that it

Shifts in several other aspects of economic structure could be added (in the structure of consumption, in the relative shares of domestic and foreign supplies, etc.). Fourth, the closely related and extremely important structures of society and its ideology have also changed rapidly. Urbanization and secularization come easily to mind as components of what sociologists term the process of modernization. Fifth, the economically developed countries, by means of the increased power of technology, particularly in transport and communication (both peaceful and warlike), have the propensity to reach out to the rest of the world—thus making for one world in the sense in which this was not true in any premodern epoch.[6] Sixth, the spread of modern economic growth, despite its worldwide partial effects, is limited in that the economic performance in countries accounting for three-quarters of world population still falls far short of the minimum levels feasible with the potential of modern technology.[7]

This brief summary of two quantitative characteristics of modern economic growth that relate to aggregate rates, two that relate to structural transformation, and two that relate to international spread, supports our working assumption that modern economic growth marks a distinct economic epoch. If the rates of aggregate growth and the speed of structural transformation in the economic, institutional, and perhaps even in the ideological, framework are so much higher than in the past as to represent a revolutionary acceleration, and if the various regions of the world are for the first time in history so closely interrelated as to be one, some new major growth source, some new epochal innovation, must have generated these radically different patterns. And one may argue that this source is the emergence of modern science as the basis of advancing technology—a breakthrough in the evolution of science that produced a potential for technology far greater than existed previously.

Yet modern growth continues many older trends, if in greatly accelerated form. This continuity is important particularly when we find that, except for Japan and possibly Russia, all presently developed countries were well in advance of the rest of the world before their modern growth and industrialization began, enjoying a comparative advantage produced by premodern trends. It is also important because it emphasizes that distinction among economic epochs is a complicated intellectual choice and that the continuation of past trends and their changing patterns over time are subjects deserving the closest attention. Does the acceleration in growth of product and productivity in many developed countries in the last two decades reflect a major change in the potential provided by science-oriented technology, or a major change in the capacity of societies to catch up with that potential? Is it a way of recouping the loss in standing, relative to such a leader as the United States, that was incurred during the depression of the 1930's and World War II? Or, finally, is it merely a reflection of the temporarily favorable climate of the *U.S.* international policies? Is the expansion into space a continuation of the old trend

took centuries for the share of the agricultural sector in the labor force to decline to 50 percent in any sizable country (i.e., excluding small "city enclaves"), a drop of 30 to 40 percentage points in the course of a single century is a strikingly fast structural change.

[6] The outward expansion of developed countries, with their European origin, goes back to long before modern economic growth, indeed, back to the Crusades. But the much augmented transportation and communication power of developed countries in the nineteenth century permitted a much greater and more direct political dominance over the colonies, the "opening up" of previously closed areas (such as Japan), and the "partition" of previously undivided areas (such as sub-Saharan Africa).

[7] For further discussion see Section IV below, which deals with the less developed countries.

of reaching out by the developed countries, or is it a precursor of a new economic epoch? These questions are clearly illustrative, but they hint at broader analytical problems suggested by the observation of modern economic growth as a distinct epoch.

The six characteristics noted are interrelated, and the interrelations among them are most significant. With the rather stable ratio of labor force to total population, a high rate of increase in per capita product means a high rate of increase in product per worker; and, with average hours of work declining, it means still higher growth rates in product per man-hour. Even if we allow for the impressive accumulation of capital, in its widest sense, the growth rate of productivity is high, and, indeed, mirrors the great rise in per capita product and in per capita pure consumption. Since the latter reflects the realized effects of advancing technology, rapid changes in production structure are inevitable—given the differential impact of technological innovations on the several production sectors, the differing income elasticity of domestic demand for various consumer goods, and the changing comparative advantage in foreign trade. As already indicated, advancing technology changes the scale of production plants and the character of the economic enterprise units. Consequently, effective participation in the modern economic system by the labor force necessitates rapid changes in its location and structure, in the relations among occupational status groups, and even in the relations between labor force and total population (the last, however, within narrow overall limits). Thus, not only are high aggregate growth rates associated with rapid changes in economic structure, but the latter are also associated with rapid changes in other aspects of society—in family formation, in urbanization, in man's views on his role and the measure of his achievement in society. The dynamic drives of modern economic growth, in the countries that entered the process ahead of others, meant a reaching out geographically; and the sequential spread of the process, facilitated by major changes in transport and communication, meant a continuous expansion to the less developed areas. At the same time, the difficulty of making the institutional and ideological transformations needed to convert the new large potential of modern technology into economic growth in the relatively short period since the late eighteenth century limited the spread of the system. Moreover, obstacles to such transformation were, and still are being, imposed on the less developed regions by the policies of the developed countries.

If the characteristics of modern economic growth are interrelated, in that one induces another in a cause and effect sequence or all are concurrent effects of a common set of underlying factors, another plausible and significant link should be noted Mass application of technological innovations, which constitutes much of the distinctive substance of modern economic growth, is closely connected with the further progress of science, in its turn the basis for additional advance in technology. While this topic is still to be studied in depth, it seems fairly clear that mass-uses of technical innovations (many based on recent scientific discoveries) provide a positive feedback. Not only do they provide a larger economic surplus for basic and applied research with long time leads and heavy capital demands, but, more specifically, they permit the development of new efficient tools for scientific use and supply new data on the behavior of natural processes under the stress of modification in economic production. In other words, many production plants in developed countries can be viewed as laboratories for the exploration of natural processes and as

centers of research on new tools, both of which are of immense service to basic and applied research in science and technology. It is no accident that the last two centuries were also periods of enormous acceleration in the contribution to the stock of useful knowledge by basic and applied research—which provided additional stimuli to new technological innovations. Thus, modern economic growth reflects an interrelation that sustains the high rate of advance through the feedback from mass applications to further knowledge. And unless some obstacles intervene, it provides a mechanism for self-sustaining technological advance, to which, given the wide expanse of the universe (relative to mankind on this planet), there are no obvious proximate limits.

III. Some Implications[8]

I turn now to a brief discussion of some social implications, of some effects of modern economic growth on conditions of life of various population groups in the countries affected. Many of these effects are of particular interest, because they are not reflected in the current measures of economic growth; and the increasing realization of this shortcoming of the measures has stimulated lively discussion of the limits and limitations of economic measurement of economic growth.

The effects on conditions of life stem partly from the major role of technological innovations in modern economic growth, and partly from the rapid shifts in the underlying production structure. To begin with the latter, the major effects of which, for example, urbanization, internal migration, shift to employee status and what might be called the merit basis of job choice, have already been noted as characteristics of modern economic growth. Two important groups of effects of this rapid transformation of economic structure deserve explicit reference.

First, the changes in conditions of life suggested by "urbanization" clearly involved a variety of costs and returns that are not now included in economic measurement, and some of which may never be susceptible to measurement. Internal migration, from the countryside to the cities (within a country, and often international) represented substantial costs in the pulling up of roots and the adjustment to the anonymity and higher costs of urban living. The learning of new skills and the declining value of previously acquired skills was clearly a costly process—to both the individuals and to society. But if such costs were omitted from measurement, as they still are in conventional accounts, so were some returns. Urban life, with its denser population, provided amenities and spiritual goods that were not available in the "dull and brutish" life of the countryside; and the new skills, once learned, were often a more adequate basis for a richer life than the old. This comment on the hidden costs and returns involved in the shift toward urban life may apply to many other costs and returns involved in other shifts imposed by economic growth, for example, in the character of participation in economic activity, in the social values, and in the new pressures on deviant members of society.

The second intriguing aspect of structural change is that it represents shifts in the relative shares in the economy of the specific population groups attached to particular production sectors. Since economic engagement represents a dominant influence in the life of people, the shift in the share of a specific sector, with its distinctive characteristics and even mode of

[8] Many of the points touched upon in this section are discussed in greater dètail in Kuznets (1971), particularly in ch. 2, pp. 75–98, which deals with the nonconventional costs of economic growth, and ch. 7, pp. 314–54, which deals with various interrelations between aggregate change and structural shifts in economic and other aspects of social structure.

life, affects the population group engaged in it. Economic growth perforce brings about a decline in the relative position of one group after another—of farmers, of small scale producers, of landowners—a change not easily accepted, and, in fact, as history teaches us, often resisted. The continuous disturbance of preexisting *relative* position of the several economic groups is pregnant with conflict—despite the rises in absolute income or product common to all groups. In some cases, these conflicts did break out into overt civil war, the Civil War in the United States being a conspicuous example. Other examples, in the early periods of industrialization among the currently developed countries, or, for that matter, more recently within some less developed countries, are not lacking.

Only if such conflicts are resolved without excessive costs, and certainly without a long-term weakening of the political fabric of the society, is modern economic growth possible. The sovereign state, with authority based on loyalty and on a community of feeling—in short, the modern national state—plays a crucial role in peacefully resolving such growth-induced conflicts. But this and other services of the national state may be costly in various ways, of which intensified nationalism is one and other effects are too familiar to mention. The records of many developed countries reveal examples of resolutions of growth conflicts, of payments for overcoming resistance and obstacles to growth, that left burdensome heritages for the following generations (notably in Germany and Japan). Of course, this is not the only economic function of the state; it can also stimulate growth and structural change. And, to mention a closely related service, it can referee, select, or discard legal and institutional innovations that are proposed in the attempt to organize and channel effectively the new production potentialities. This, too, is a matter that may generate conflicts, since different legal and institutional arrangements may have different effects on the several economic groups in society.

In that modern economic growth has to contend with the resolution of incipient conflicts continuously generated by rapid changes in economic and social structure, it may be described as a process of controlled revolution. The succession of technological innovations characteristic of modern economic growth and the social innovations that provide the needed adjustments are major factors affecting economic and social structure. But these innovations have other effects that deserve explicit mention; and while they are discussed below in terms of effects of technological innovations, the conclusions apply *pari passu* to innovations in legal forms, in institutional structure, and even in ideology.

A technological innovation, particularly one based on a recent major invention, represents a venture into the partly unknown, something not fully known until the mass spread of the innovation reveals the full range of direct and related effects. An invention is a major one if it provides the basis for extensive applications and improvements (for example, the stationary steam engine in the form attributable mostly to James Watt). Its cumulative effects, all new, extend over a long period and result in an enormous transformation of economic production and of production relations. But these new effects can hardly be fully anticipated or properly evaluated in advance (and sometimes not even post facto). This is true also of electric power, the internal combustion engine, atomic energy, the application of short rays to communication and computation, the inventions resulting in such new industrial materials as steel, aluminum, and plastics, and so on through a long list that marked modern economic growth. Even when the

technological innovation is an adaptation of a known technique by a follower country, the results may not be fully foreseeable, for they represent the combination of something known, the technology, with something new, an institutional and ideological framework with which it has not previously been combined. Needless to say, the element of the uniquely new, of exploration into the unknown, was also prominent in premodern times, since innovations in knowledge and technology are the prerequisites for any significant growth. But the *rate* of succession of such innovations was clearly more rapid in modern economic growth, and provided the base for a higher rate of aggregate growth.

The effects of such ventures into the new and partly unknown are numerous. Those of most interest here are the *surprises*, the unexpected results, which may be positive or negative. An invention or innovation may prove far more productive, and induce a far wider mass application and many more cumulative improvements than were dreamed of by the inventor and the pioneer group of entrepreneurs. Or the mass application of a major invention may produce unexpected diseconomies of a scale that could hardly be foreseen in the early phases of its diffusion. Examples of both positive and negative surprises abound. Many Schumpeterian entrepreneurs failed to grasp, by a wide margin, the full scope and significance of the innovations that they were promoting and that eventually brought them fame and fortune. And most of us can point at the unexpected negative effects of some technological or social invention that first appeared to be an unlimited blessing.

The significant aspect here is that the surprises cannot be viewed as accidents: they are inherent in the process of technological (and social) innovation in that it contains an element of the unknown. Furthermore, the diffusion of a major innovation is a long and complicated sequence that cannot be accurately forecast, with an initial economic effect that may generate responses in other processes. These will, in turn, change the conditions under which the innovation exercises its effect on human welfare, and raise further problems of adjustment. To illustrate: we can today follow easily the sequence from the introduction of the passenger car as a mass means of transportation, to the growth of the suburbs, to the movement of the more affluent from the city centers, to the concentration of lower income recipients and unemployed immigrants in the slums of the inner city core, to the acute urban problems, financial and other, and to the trend toward metropolitan consolidation. But the nature and implications of this sequence were certainly not apparent in the 1920's, when passenger cars began their mass service function in the United States.

Indeed, to push this speculative line further, one can argue that all economic growth brings *some* unexpected results in its wake, positive as well as negative, with the latter taking on greater importance as the mass effects of major innovations are felt and the needs that they are meant to satisfy are met. If the argument is valid, modern economic growth, with the rapid succession of innovations and shortening period of their mass diffusion, must be accompanied by a relatively high incidence of negative effects. Yet one must not forget that premodern economic growth had similar problems, which, with the weaker technology, may have loomed even larger. Even if we disregard the threatening exhaustion of natural resources, a problem that so concerned Classical (and implicitly even Marxian) economics, and consider only early urbanization, one major negative effect was the significant rise in death rates as population moved from the more salubrious countryside to the infection-

prone denser conditions of unsanitary cities. Two points are relevant here. First, the negative effects of growth have never been viewed as so far outweighing its positive contribution as to lead to its renunciation—no matter how crude the underlying calculus may have been. Second, one may assume that once an unexpected negative result of growth emerges, the potential of material and social technology is aimed at its reduction or removal. In many cases these negative results were allowed to accumulate and to become serious technological or social problems because it was so difficult to foresee them early enough in the process to take effective preventive or ameliorative action. Even when such action was initiated, there may have been delay in the effective technological or policy solution. Still, one may justifiably argue, in the light of the history of economic growth, in which a succession of such unexpected negative results has been overcome, that any specific problem so generated will be temporary—although we shall never be free of them, no matter what economic development is attained.

IV. The Less Developed Countries

Two major groups of factors appear to have limited the spread of modern economic growth. First, as already suggested, such growth demands a stable, but flexible, political and social framework, capable of accommodating rapid structural change and resolving the conflicts that it generates, while encouraging the growth-promoting groups in society. Such a framework is not easily or rapidly attained, as evidenced by the long struggles toward it even in some of the presently developed countries in the nineteenth and early twentieth centuries. Japan is the only nation outside of those rooted in European civilization that has joined the group of developed countries so far. Emergence of a modern framework for economic growth may be especially difficult if it involves elements peculiar to European civilization for which substitutes are not easily found. Second, the increasingly national cast of organization in developed countries made for policies toward other parts of the world that, while introducing some modern economic and social elements, were, in many areas, clearly inhibiting. These policies ranged from the imposition of colonial status to other limitations on political freedom, and, as a result, political independence and removal of the inferior status of the native members of the community, rather than economic advance, were given top priority.

Whatever the weight of the several factors in explaining the failure of the less developed countries to take advantage of the potential of modern economic growth, a topic that, in its range from imperialist exploitation to backwardness of the native economic and social framework, lends itself to passionate and biased polemic, the factual findings are clear. At present, about two-thirds or more of world population is in the economically less developed group. Even more significant is the concentration of the population at the low end of the product per capita range. In 1965, the last year for which we have worldwide comparable product estimates, the per capita *GDP* (at market prices) of 1.72 billion out of a world total of 3.27 billion, was less than \$120, whereas 0.86 billion in economically developed countries had a per capita product of some \$1900. Even with this narrow definition of less developed countries, the intermediate group was less than 0.7 billion, or less than 20 percent of world population.[9] The preponderant

[9] The underlying data are from Everett Hagen and Oli Hawrylyshyn. These are primarily from United Nations publications, supplemented by some auxiliary sources (mostly for the Communist countries), and use conventional conversion rates to *U.S.* dollars in 1965. The estimates for the Communist countries have been adjusted to conform to the international *GDP* concept.

population was thus divided between the very low and the rather high level of per capita economic performance. Obviously, this aspect of modern economic growth deserves our greatest attention, and the fact that the quantitative data and our knowledge of the institutional structures of the less developed countries are, at the moment, far more limited than our knowledge of the developed areas, is not reason enough for us to ignore it.

Several preliminary findings, or rather plausible impressions, may be noted. First, the group of less developed countries, particularly if we widen it (as we should) to include those with a per capita product somewhat larger than $120 (in 1965 prices), covers an extremely wide range in size, in the relations between population and natural resources, in major inherited institutions, and in the past impact upon them of the developed countries (coming as it did at different times and from different sources). There is a striking contrast, for example, in terms of population size, between the giants like Mainland China and India, on the one hand, and the scores of tiny states in Africa and Latin America; as there is between the timing of direct Western impact on Africa and of that on many countries in Latin America. Furthermore, the remarkable institutions by which the Sinic and East Indian civilizations produced the unified, huge societies that dwarfed in size any that originated in Europe until recently. bore little resemblance to those that structured the American Indian societies or those that fashioned the numerous tribal societies of Africa.

Generalizations about less developed countries must be carefully and critically scrutinized in the light of this wide variety of conditions and institutions. To be sure, their common failure to exploit the potential of modern economic growth means several specific common features: a low per capita product, a large share of agriculture or other extractive industries, a generally small scale of production. But the specific parameters differ widely, and because the obstacles to growth may differ critically in their substance, they may suggest different policy directions.

Second, the growth position of the less developed countries today is significantly different, in many respects, from that of the presently developed countries on the eve of their entry into modern economic growth (with the possible exception of Japan, and one cannot be sure even of that). The less developed areas that account for the largest part of the world population today are at much lower per capita product levels than were the developed countries just before their industrialization; and the latter at that time were economically in advance of the rest of the world, not at the low end of the per capita product range. The very magnitudes, as well as some of the basic conditions, are quite different: no country that entered modern economic growth (except Russia) approached the size of India or China, or even of Pakistan and Indonesia; and no currently developed country had to adjust to the very high rates of natural increase of population that have characterized many less developed countries over the last two or three decades. Particularly before World War I, the older European countries, and to some extent even Japan, relieved some strains of industrialization by substantial emigration of the displaced population to areas with more favorable opportunities—an avenue closed to the populous less developed countries today. Of course, the stock of material and social technology that can be tapped by less developed countries today is enormously

The developed countries include most countries with per capita *GDP* of $1000 or more and Japan, but exclude those small countries with a high *GDP* per capita that is due to exceptional natural endowments (for example, Netherlands Antilles, Puerto Rico, Kuwait, and Qatar).

larger than that available in the nineteenth and even early twentieth centuries. But it is precisely this combination of greater backwardness and seemingly greater backlog of technology that makes for the significant differences between the growth position of the less developed countries today and that of the developed countries when they were entering the modern economic growth process.

Finally, it may well be that, despite the tremendous accumulation of material and social technology, the stock of innovations most suitable to the needs of the less developed countries is not too abundant. Even if one were to argue that progress in basic science may not be closely tied to the technological needs of the country of origin (and even that may be disputed), unquestionably the applied advances, the inventions and tools, are a response to the specific needs of the country within which they originate. This was certainly true of several major inventions associated with the Industrial Revolution in England, and illustrations abound of necessity as the mother of invention. To the extent that this is true, and that the conditions of production in the developed countries differed greatly from those in the populous less developed countries today, the material technology evolved in the developed countries may not supply the needed innovations. Nor is the social technology that evolved in the developed countries likely to provide models of institutions or arrangements suitable to the diverse institutional and population-size backgrounds of many less developed countries. Thus, modern technology with its emphasis on labor-saving inventions may not be suited to countries with a plethora of labor but a scarcity of other factors, such as land and water; and modern institutions, with their emphasis on personal responsibility and pursuit of economic interest, may not be suited to the more traditional life patterns of the agricultural communities that predominate in many less developed countries. These comments should not be interpreted as denying the value of many transferable parts of modern technology; they are merely intended to stress the possible shortage of material and social tools specifically fitted to the different needs of the less developed countries.

If the observations just made are valid, several implications for the growth problems of the less developed countries follow. I hesitate to formulate them explicitly since the data and the stock of knowledge on which the observations rest are limited. But at least one implication is sufficiently intriguing, and seems to be illuminating of many recent events in the field, to warrant a brief note. It is that a substantial economic advance in the less developed countries may require modifications in the available stock of material technology, and probably even greater innovations in political and social structure. It will not be a matter of merely borrowing existing tools, material and social; or of directly applying past patterns of growth, merely allowing for the difference in parameters.

The innovational requirements are likely to be particularly great in the social and political structures. The rather violent changes in these structures that occurred in those countries that have forged ahead with highly forced industrialization under Communist auspices, the pioneer entry going back over forty years (beginning with the first Five-Year Plan in the *USSR*), are conspicuous illustrations of the kind of social invention and innovation that may be involved. And the variants even of Communist organization, let alone those of democracy and of non-Communist authoritarianism, are familiar. It would be an oversimplification to argue that these innovations in the social and political structures were made primarily in response to the strain between economic backwardness and the potential of modern economic growth; or to claim that they were inexor-

able effects of antecedent history. But to whatever the struggle for political and social organization is a response, once it has been resolved, the results shape significantly the conditions under which economic growth can occur. It seems highly probable that a long period of experimentation and struggle toward a viable political framework compatible with adequate economic growth lies ahead for most less developed countries of today; and this process will become more intensive and acute as the *perceived* gap widens between what has been attained and what is attainable with modern economic growth. While an economist can argue that some aspects of growth must be present because they are indispensable components (i.e., industrialization, large scale of production, etc.), even their parameters are bound to be variable; and many specific characteristics will be so dependent upon the outcome of the social and political innovations that extrapolation from the past is extremely hazardous.

V. Concluding Comments

The aim of the discussion was to sketch the major characteristics of modern economic growth, and to note some of the implications that the empirical study of economic growth of nations suggests. This study goes back to the beginning of our discipline, as indicated by the title of Adam Smith's founding treatise, *Wealth of Nations*, which could as well have been called the Economic Growth of Nations. But the quantitative base and interest in economic growth have widened greatly in the last three to four decades, and the accumulated results of past study of economic history and of past economic analysis could be combined with the richer stock of quantitative data to advance the empirical study of the process. The sketch above draws upon the results of many and widely varied studies in many countries, most of them economically developed; and the discussion reflects a wide collective effort, however individual some of my interpretations may be.

The most distinctive feature of modern economic growth is the combination of a high rate of aggregate growth with disrupting effects and new "problems." The high rate of growth is sustained by the interplay between mass applications of technological innovations based on additions to the stock of knowledge and further additions to that stock. The disrupting effects are those imposed by the rapid rate of change in economic and social structure. The problems are the unexpected and unforeseeable results of the spread of innovations (with emphasis on the new and unknown indicated by that term). Added to this is the range of problems raised by the slow spread of economic growth to the less developed countries, all of which have a long history, separate and relatively isolated from the areas within which modern economic growth originated. Thus, concurrent with the remarkable positive achievements of modern economic growth are unexpected negative results even within the developed countries; while the less developed countries are struggling in the attempt to use the large potential of modern technology in order to assume an adequate role in the one and interdependent world (from which they cannot withdraw even if they wished to do so).

We have stressed the problem aspects of modern economic growth because they indicate the directions of further research in the field. These aspects, the "surprises" and the implicit explanatory "puzzles," are problems not only in the sense of departures from the desirable (that may call for policy amelioration) but also in the sense that our quantitative data and particularly our analytical hypotheses do not provide us with a full view and explanation. As already noted, the conventional measures of national product and its components do not reflect many costs of ad-

justment in the economic and social structures to the channeling of major technological innovations; and, indeed, also omit some positive returns. The earlier theory that underlies these measures defined the productive factors in a relatively narrow way, and left the rise in productivity as an unexplained gap, as a measure of our ignorance. This shortcoming of the theory in confrontation with the new findings, has led to a lively discussion in the field in recent years, and to attempts to expand the national accounting framework to encompass the so far hidden but clearly important costs, for example, in education as capital investment, in the shift to urban life, or in the pollution and other negative results of mass production. These efforts will also uncover some so far unmeasured positive returns—in the way of greater health and longevity, greater mobility, more leisure, less income inequality, and the like. The related efforts to include the additions to knowledge in the framework of economic analysis, the greater attention to the uses of time and to the household as the focus of economic decision not only on consumption but also on investment, are steps in the same direction. It seems fairly clear that a number of analytical and measurement problems remain in the theory and in the evaluation of economic growth in the developed countries themselves; and that one may look forward to major changes in some aspects of the analysis, in national economic accounting, and in the stock of empirical findings, which will occupy economists in the developed countries in the years ahead.

For the less developed countries the tasks of economic research are somewhat different: the great need is for a wider supply of tested data, which means essentially data that have been scrutinized in the process of use for economic analysis. As already noted, the stock of data and of economic analysis is far poorer for these countries than that for the developed countries—a parallel to the smaller relative supply of material capital. Yet in recent years there has been rapid accumulation of data for many less developed areas, other than those that, like Mainland China, view data as information useful to their enemies (external or internal) and are therefore either not revealed by government or possibly not even collected. The lag has been in the analysis of these data by economists and other social science scholars, because of the scarcity of such scholars who cannot be spared for research within the less developed countries themselves and because of the natural preoccupation of economists in the developed countries with the problems of their own countries. One may hope, but with limited expectations, that the task of refining analysis and measurement in the developed countries will not be pursued to the exclusion or neglect of badly needed studies of the less developed countries, studies that would deal with the quantitative bases and institutional conditions of their performance, in addition to those concentrating on what appear to be their major bottlenecks and the seemingly optimal policy prescriptions.

REFERENCES

P. Bairoch et al., *The Working Population and Its Structure, International Historical Statistics*, vol. I, Brussels 1968.

E. E. Hagen and O. Hawrylyshyn, "Analysis of World Income and Growth, 1955–1965," *Econ. Develop. Cult. Change*, Oct. 1969, *18*, Part II, 1–96.

S. Kuznets, *Economic Growth of Nations: Total Output and Production Structure*, Cambridge, Mass. 1971.

———, *Modern Economic Growth: Rate, Structure, and Spread*, New Haven 1966.

United Nations, *Yearbook of National Accounts Statistics, 1969*, vol. II, *International Tables*, New York 1970.

The Economics of the Network-Affiliate Relationship in the Television Broadcasting Industry

By Stanley M. Besen and Ronald Soligo*

Under various acts of Congress, the Federal Communications Commission is empowered to regulate the television broadcasting industry in the public interest.[1] Among the Commission's objectives have been the promotion of program diversity, the encouragement of local programming, and the maintenance of widespread ownership of broadcasting facilities.[2] In order to further these objectives, the FCC has limited the number of local broadcasting stations which can be owned by any one group of individuals or corporations, including networks, and forbidden network affiliates from entering into agreements to accept specific amounts of network programming during specified hours. In addition, the FCC has recently established a ceiling on the number of hours during prime time which a local station can "clear" for network programs.[3] In this paper we have formulated a model of the network-affiliate relationship which allows us to examine the effects of each of these regulations and to examine the extent to which FCC objectives are consistent with one another.

* Associate professors of economics, Rice University. We wish to acknowledge the helpful comments of Marian Krzyzaniak, Harold Barnett, and Edward Greenberg on earlier drafts.

[1] For a discussion of the development of government regulation of broadcasting, see Hiram Jome and Ronald Coase.

[2] For a discussion of FCC objectives, see Hyman Goldin.

[3] In the jargon of the industry, to clear time for the network means to accept network programs; i.e., to keep the time slot clear of other programs.

I. The Network-Affiliate Contract

The specific details of the typical affiliation agreement have been documented by John Peterman and by Harlan Blake and Jack Blum. For our purposes, the important features of these contracts are the methods by which advertising revenue is shared by the network and affiliate. The standard agreement separates broadcasting time into two categories: announcement time and program time. The announcement time is the brief interval between programs in which the station identifies itself and sells spot commercials. Local stations usually keep all revenue from these sales. Program time is the time during which program material is being broadcast. The networks sell segments of this time for commercial messages. The price which the network charges for this time is determined separately for each station carrying the network program and the revenue which the network collects from these sales is shared with the local affiliates except for a certain specified amount of cleared time for which the network retains all of the revenue.

The important dimensions of the affiliation agreement are i) the length of announcement time allowed by the network, ii) the proportion of program time revenue returned to affiliates, and iii) the number of hours for which all program time revenue is kept by the network. (Hereafter this is referred to as uncompensated time.)

II. A Simple Model of Affiliate Behavior

In order to understand the basic features of the process by which the terms of the affiliation agreement, the number of affiliates, and the number of hours cleared are determined, we develop a model of affiliate behavior. Initially, we consider a highly simplified version of the model. As we proceed, we relax a number of its restrictive assumptions. In the simplified model we assume:

1. The proportion of program time revenue, p, returned to a station is assumed to be the same for all programs cleared regardless of the time during the day at which it is cleared.
2. The amount of announcement time is fixed and not subject to change by the network or station.
3. There is no uncompensated time.
4. There are no constraints on the value or number of hours which the local station must or must not clear for network programming. The affiliates can choose to clear anything from zero to 100 percent of its time for the networks.
5. All network programming costs are fixed. The marginal cost of supplying an additional hour of existing network programming to an affiliate or potential affiliate is zero.
6. In making an agreement, both the station and the network act as if the behavior of other stations and networks is not affected by the nature of the agreement reached.

Since different hours of the day or week will have different values and hence one hour is qualitatively not the same as another, we take as the decision variable for a particular station the value of the time cleared for network programming. The total value of hours cleared is equal to the sum of the amounts which each hour cleared will yield in program time advertising revenue. This in turn is affected by the size of the market (population, number of TV sets, number of TV stations, time of day, what other stations are broadcasting and so on).

The value of hours cleared by a station is, given all of the above assumptions and parameters, a function of p, the proportion of the program time revenue which is returned to the station. Given the value of announcement time if it adjoins a local program, its value if it adjoins a network program, the value of program time for network and nonnetwork programs, and the cost of nonnetwork programs, it is possible to array time periods according to the value of p for which it will pay the local station to clear the network program.

If A_i^L represents the value of advertising during program time for a nonnetwork program during the ith hour of the broadcasting day, C_i the cost of the program, and S_i^L the value of advertising during the announcement time adjoining the program, the return to the local station from a nonnetwork program during that time period is

$$V_i^L = A_i^L + S_i^L - C_i \tag{1}$$

If A_i^N is the value of program time for a network program during the ith hour and S_i^N, the value of the announcement time adjoining the network program, the return to the affiliate from clearing the network program is equal to $pA_i^N+S_i^N$. In deciding whether or not to clear the ith hour for a network program, the affiliate compares these two returns. If $pA_i^N+S_i^N>V_i^L$ the station accepts the network program. Otherwise it does not.

For each hour of broadcasting time the affiliate can calculate the value of p for which it would be indifferent between network and nonnetwork programming. If we denote this p by $\hat{p}_i$ we have, for each i

$$\hat{p}_i A_i^N + S_i^N = V_i^L \tag{2}$$

or

$$\hat{p}_i A_i^N = V_i^L - S_i^N \tag{3}$$

If we array program hours according to $\hat{p}_i$, we can construct a schedule which indicates, for each level of p, the value of hours which the affiliate would be willing to clear for network programming.[4] The higher the p, the greater will be both the number and the value of hours cleared for network programs.[5]

In Figure 1, we have plotted such a "supply" curve for a representative station as BC. The horizontal axis measures the total amount advertisers pay for program time commercial messages during the hours cleared for network programs. The point A^* represents the acceptance of all network programs. We have drawn the curve so that it has a negative intercept. In general, we expect that there are some hours for which $S_i^N > V_i^L$, in which case the local affiliate would even be willing to pay the network for the right to clear network programs, i.e., the value of $\hat{p}$ would be negative. This situation arises because the value of announcement time is greater if the time is adjacent to a network program than if it adjoins a locally originated program and because by clearing a network program a station avoids the cost of program origination.[6]

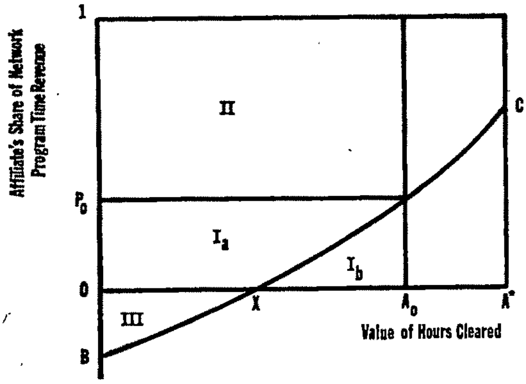


FIGURE 1

If the level of p set by the network is p_O, the station will clear hours worth A_O. The revenue which it collects from the network as its share of the network program time revenues is given by the rectangular area $I_a + I_b$. The network's revenue from the programs cleared on this station is area II. The net profit to the station from network affiliation, i.e., revenues over and above profits, if any, which could be earned from locally originated programs for these hours, is given by the area $I_a + III$.[7] This is the value of network affiliation to the local station. It is a quasi rent which accrues to the local affiliate so long as the network does not engage in price discrimination by establishing a different p for each hour of time cleared.

What value of p will be set by the network for the affiliate depicted in Figure 1? As the network raises p, the value of hours

[4] Note also that we are assuming that, for any i, V_i^L is independent of the amount of time cleared for network programming. This assumption is undoubtedly not strictly correct.

[5] One alternative available to a station is to originate local programs. Another is the purchase of a program from a company specializing in the production of programs. In this case, too, the local station retains all revenue from program advertising. Still another alternative is the so-called barter arrangement. Here, an advertiser supplies a program free of charge to a local station and the latter is permitted to sell some program advertising time from which it can retain all revenues. Even if less advertising revenue accrues to the local station under this arrangement because the supplier of the program does not pay for his advertising time, it may still be the best alternative to a network program since the station avoids the cost of originating or purchasing a program. Of course, it is the return from the best alternative which will be compared to the gains from accepting network programming.

[6] There may be markets where local programming is very profitable, relative to network programs. In this case, the local station will not be willing to clear any network programs without a positive p. The supply curve for this station would have a positive intercept. These cases are, however, rare. There are no stations which do not affiliate with a network if they can.

[7] Recall that the supply curve represents the opportunity cost of program time to the affiliate. Hours to the left of X have negative opportunity cost; i.e., the affiliate would pay for network programming if necessary.

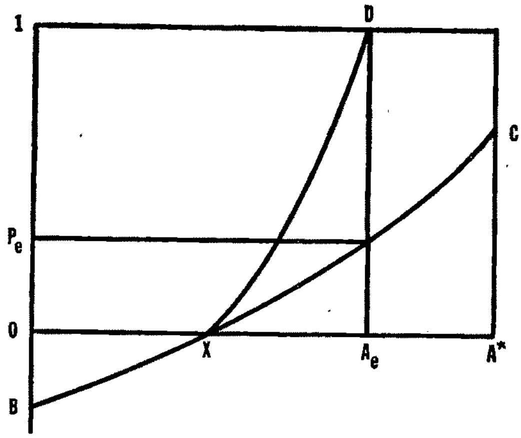


FIGURE 2

cleared by the affiliate increases. The profits that the network earns on the newly cleared units is at least partially offset, however, by the decline in revenue on previously cleared units. In other words, since a lower value of $1-p$ must be accepted on all previously cleared units, the return from an additional unit cleared is less than $1-p$ for that unit. The network will, therefore, employ the curve marginal to the supply curve, XD in Figure 2, in determining the value of p for which its profits are a maximum.[8] By assumption, the marginal cost of supplying a program to an additional station is zero.[9] Therefore, profits are maximized when the curve marginal to the supply curve intersects the horizontal line for $p=1$. Given this, the profit maximizing value of p can be determined.[10] In Figure 2 this is p_e. In equilibrium, the affiliated station is indifferent between clearing the last hour of network programming, A_e, and not doing so. Note that the price charged by the network, $(1-p)$ times advertising revenue, exceeds the cost (zero by assumption) of supplying an additional hour of programming to the affiliate.

III. Price Discrimination by Network

Suppose that the network has considerable market power since the number of stations exceeds the number of networks in a given market and some stations will not be able to obtain an affiliation. In principle, the network should be able to extract the entire increment to the profits of an affiliate generated by network affiliation. One way to do this would be to charge different prices for each program cleared.[11] Another way of accomplishing the same objective is not to compensate the affiliate for all programs cleared and to require in the contract of affiliation that the affiliate clear some minimum number of hours. We consider these, in turn.

Uncompensated Time

The uncompensated time constraint can be described as follows: If the value of advertising during an hour of prime time on a network program is A_P^N and if the constraint is H hours,[12] then the affiliate does not share in program time advertising revenue on the first HA_P^N dollars worth of network time that it clears.

Consider an affiliate which was clearing time periods of value greater than HA_P^N in the absence of the constraint. When a constraint of that amount is imposed, the affiliate will clearly wish to satisfy it with the smallest loss of profits. Given p, it will never clear hours not previously cleared

[8] The curve XD does not extend below the horizontal axis because the network will have programs of value OX cleared even if $p=0$. If the network were permitted to make p negative, i.e., to charge the affiliate for a cleared program, the curve marginal to the supply curve would begin at B.

[9] This departs from reality in that the network typically pays the AT&T interconnection charge.

[10] For example, if the supply curve were a straight line, p would equal 1/2.

[11] Notice that this would require a negative p for programs to the left of X. We will return to this point shortly.

[12] The constraint is specified, in the affiliation contract, in terms of prime time hours or equivalents thereof but is, in effect, a claim to an amount of revenue rather than to a number of hours.

(when there was no constraint) in order to satisfy the constraint. In Figure 3, hours previously cleared have an aggregate value of A_O for $p=p_O$. The constraint is shown as $\bar{A}$. If, after the imposition of the constraint, the affiliate continues to clear time periods of value greater than $\bar{A}$ it will continue to clear A_O since eliminating a single program does not affect the constraint and contributes a positive amount, $(pA^N+S^N-V^L)$, to profits. But if area I is greater than area II, the affiliate would be better off if it reduced the value of hours cleared to OX. This is a preferred position because the gain from clearing hours between $\bar{A}$ and A_O is less than the foregone income on hours between X and $\bar{A}$. The affiliate would continue to clear hours up to X, however, since all hours to the left of X have the characteristic that S^N exceeds V^L and would be cleared even if p were negative.

What effect does the possibility of employing an uncompensated time constraint have on the policy pursued by a network with respect to a particular station? It is easy to see that the optimal policy for the network is to set p sufficiently high so that all network programs are cleared and to adjust the constraint so that the affiliate's profits from network affiliation are almost

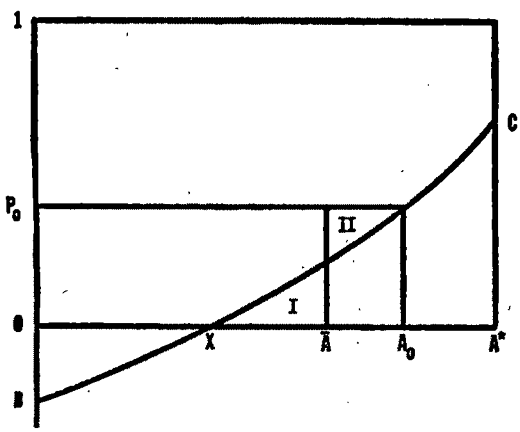


FIGURE 3

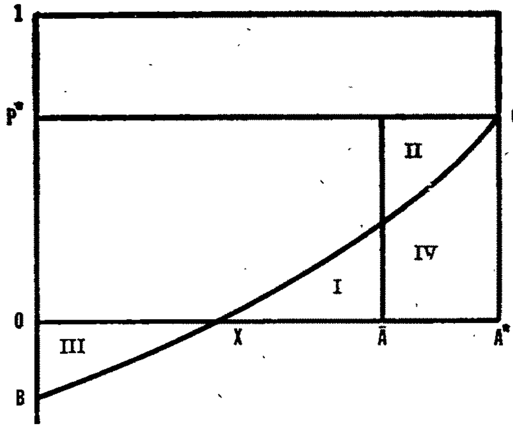


FIGURE 4

eliminated. This is demonstrated in Figure 4, where A^* represents the acceptance of all network programs, the sharing ratio is set at p^* and the constraint at $\bar{A}$ so that area I is equal to area II. If the constraint were any larger, the affiliate would be induced to reduce the number of hours cleared to OX. Notice that in this case the affiliate would continue to earn profits from network affiliation equal to area III.

If the network acted as a perfect price discriminator and set $p=\hat{p}$ for each hour to the right of X, profits accruing to the affiliate on these hours would be equal to areas I and IV. Since these profits are equal to what the station would earn if it were an independent station, profits from network affiliation on hours to the right of X would be zero. The uncompensated time constraint thus appears as a relatively simple, yet sophisticated, device for practicing price discrimination among hours. Instead of setting prices on each program individually the network simply establishes a sharing ratio sufficiently high so that all network programs are cleared and then adjusts the constraint so as to eliminate almost all of the profits from affiliation and, at the same time, maximizes the profits of the network.

The above arrangement with a high sharing ratio and a large amount of un-

compensated time is very similar to the incentive contract initiated by CBS in 1961 and subsequently overturned by the FCC in 1962. The CBS contract called for affiliates to receive 10 percent of the revenue from the first three-fifths of the programs that CBS asked them to clear and 60 percent of the revenue from all additional programs cleared. This sharing ratio contrasts sharply with the 30 percent which prevailed both before and since the imposition of the incentive contract. The FCC overturned the contract on the grounds that it unduly restricted the ability of nonnetwork program sources to obtain clearances on CBS affiliates.[13,14]

The overturning of the CBS incentive contract raises the question of how the terms of network-affiliate contracts are affected by it. From the comments made at the time by the FCC, we deduce that the constraint imposed is in terms of a maximum sharing ratio permitted,[15] say p_{max}. With such a restriction, the network would set the constraint so that the affiliate is indifferent between clearing the number of hours cleared at p_{max} and OX, the minimum amount that affiliates would always want to clear. The situation is depicted in Figure 5 where area I is equal to area II. The effect of the constraint, and the principal one intended by FCC policy, is to

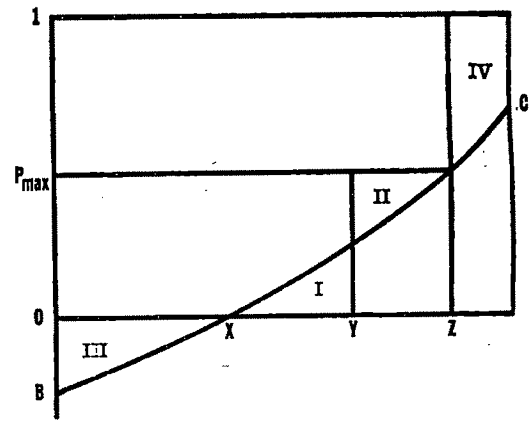


FIGURE 5

reduce the number of network hours which an affiliate will accept. The quasi rents earned by the affiliate as a result of network affiliation are still restricted to area III but the profits earned by the network are much reduced (by area IV) from those obtained where there is no restriction on p.

Option Time

A network which has the power to practice price discrimination need not settle for almost all of the quasi rent it generates for its affiliates. Under certain circumstances it may also be able to extract the quasi rent earned on hours cleared to the left of OX. The device employed is called option time[16] whereby the affiliate must clear some minimum number of hours for network programs.

Consider the situation depicted in Figure 6 which partly reproduces Figure 4. If the value of uncompensated time were increased beyond $\bar{A}$, say to $\hat{A}$, the affiliate would clear only OX hours, since with a constraint equal to $\bar{A}$ the station was indifferent between clearing OA^* hours and OX hours. The change in the constraint

[13] See *Wall Street Journal*, Apr. 13, 1962, p. 4, and June 1, 1962, p. 14.

[14] In some instances, there has been a more direct form of price discrimination by hours. For example, in 1956, the contract of WNBF affiliate in Binghamton, New York, called for a 25 percent sharing ratio on the first 20 hours cleared, a 35 percent ratio on the next 10 hours, and 40 percent on all additional hours. At the same time the CBS affiliate in Youngstown, Ohio, WKBN, received 10 percent on the first 5 hours cleared, 20 percent on the next 5, and 30 percent on all other hours. However, such sliding scale arrangements are apparently not common. On the terms of contracts see U.S. Congress, Hearings: *Monopoly Problems in Regulated Industries, Part 2, Television*.

[15] This is the aspect of the contract that limits the ability of nonnetwork program sources to compete with networks for program time.

[16] Option time was forbidden to be included in the contract of affiliation in 1963. Yet, some observers of the industry believe, local stations are often "forced" to accept programs they would otherwise not clear. The threat used against them is the withdrawal of affiliation.

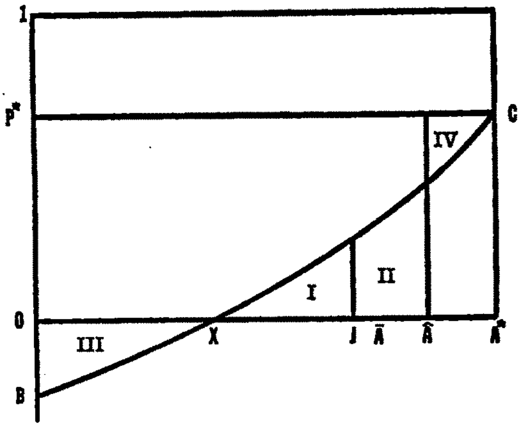


FIGURE 6

induces a shift to OX. But suppose that the contract of affiliation requires that hours of value at least OJ be cleared. Then the station must decide whether it will clear OJ hours, and perhaps some larger number of hours, or refuse the opportunity to affiliate. Which course it will choose depends on the relationship between areas III and I in Figure 6. If the network sets OJ so that these areas are equal, then at least OJ hours will be cleared. The loss on the hours between OX and OJ, area I, is just compensated for by the quasi rents earned on the first OX hours, area III. In fact, the station then considers, as a marginal decision, whether to clear some larger number of hours. If area IV is greater than area II, it will clear all network programming since the loss on hours between J and $\hat{A}$ is more than compensated for by the gain on hours between $\hat{A}$ and A^*. If $\hat{A}$ is set so that areas II and IV are equal in addition to setting the option time OJ so that areas I and III are equal, the network will be able to extract all of the quasi rents that it generates.

To summarize: The use of uncompensated time is a device for practicing price discrimination which can eliminate almost all of the quasi rents which would otherwise be received by a network affiliate as a result of affiliation. Combined with option time, or some "unofficial" requirement that some minimum number of hours be cleared, all of the quasi rents can, in principle at least, be eliminated. Placing restriction on the maximum sharing ratio results in a decline in network profits with no corresponding gain in affiliate profits. The bar against option time, if it is effective, may result in some quasi rents being retained by the affiliate.

IV. Price Discrimination by Stations

To this point we have considered situations in which the networks have considerable market power vis-à-vis potential affiliates. This presumably occurs in markets in which there are very many stations, each vying for the right to be an affiliate. In other markets, however, the situation is reversed. A small number (one or two) stations exist and each of the networks would wish to have the station as its affiliate. Here, competition among the networks should completely eliminate any profits that networks obtain from having an additional affiliate. Rent accrues to the station from being an affiliate as well as from the fact that it is one of a small number of stations in the market. If, from the point of view of the station, one of the network affiliations is more desirable than the others in the sense that advertising revenues from that network's programs are larger, then that network can obtain the portion of revenue which results from its superior programming. This is the portion of the rents which results from the differential productivity of the networks.

V. Affiliate Profits and Market Structure

As we have pointed out above, our model has some implications for the profitability of network affiliates in markets in which there are more stations than networks. In the absence of any FCC restrictions, some combination of sharing ratio, uncompensated time, and option time should succeed in eliminating all quasi rents which would

otherwise accrue to a station from its affiliation with a network. In such markets, network stations should be no more profitable than are independents. If, however, option time is not permitted, and the restriction is effective, a portion of the quasi rents will continue to accrue to affiliates.

There is some empirical evidence on this point. Edward Greenberg has shown that, in markets with four or more stations, network affiliates are more profitable than are independents when other determinants of profitability are allowed for. This indicates that, even in markets in which competition among stations for the right to a network affiliation might be expected to lead to the extraction by networks of all profits from affiliation this does not in fact occur.[17] The implication of this is that either the FCC prohibition of option time is effective or that network pricing policies are not such as to fully eliminate the additional revenue that network affiliation produces.[18] Further, it follows, a fortiori, that in markets with less than four stations, where the bargaining power of the network is weaker, it is even more likely that local stations are able to capture at least a portion of the incremental revenues which results from affiliation.

VI. The Prime Time Access Rule

The FCC has recently enacted a rule which restricts the number of prime time hours which an affiliate can clear to a network.[19] Assuming that the affected affiliates do not now satisfy this constraint, as is likely to be the case, what is its effect? We assume that the situation which the Prime Time Access Rule will affect is that depicted in Figure 5 where there is both an upper bound to the sharing ratio and the absence of option time. In Figure 5, the value of uncompensated time is set at OY hours. If the program omitted by the network as a result of the Prime Time Access Rule lies to the left of OY, the impact of the rule will be to raise the value of uncompensated time. The affiliate's loss on programming to the left of Y caused by the existence of uncompensated time is now reduced with the elimination of one program. The network can now increase uncompensated time in order to, once again, equalize areas I and II. If, on the other hand, the program omitted lies to the right of OY, the impact will be to reduce the value of uncompensated time. To predict the precise impact of the rule would, therefore, require detailed knowledge of the value to the affiliates of the programs to be omitted.[20] In either case, network profits will be reduced.[21,22]

[17] Harvey Levin also found that network affiliation is significant as a determinant of the value of a station even when the effect of other factors is allowed for.

[18] There is one further possibility: Greenberg's results may reflect the fact that UHF stations are less profitable than VHF stations, and that network affiliates are largely the latter. A UHF-VHF variable was not significant in Greenberg's regressions but this may be explained by the multicollinearity between this variable and the network affiliation variable.

[19] The rule, which would apply only to affiliates located in the 50 largest markets, would prohibit stations from carrying more than 3 hours of network programming other than news between 7 and 11 P.M. Because each offers a half hour early evening news program the practical effect would be to reduce network programs by one-half hour each night. Networks would be forbidden from extending news programming. See *Wall Street Journal*, Apr. 15, 1970, p. 2 and May 4, 1970, p. 10.

[20] The omitted program is most likely to lie to the right of OY. Since the network is assumed to vary the uncompensated time constraint so as to extract all quasi rents from network programs, the contribution to the network's profits from any one program is equal to the difference between the amount of program time revenue (net of costs) paid to affiliates and the amount which must be paid in order to induce the affiliate to carry the program. In the case where all prime time hours produce the same program time revenue (net of costs), the decrease in the network's profits due to the Prime Time Access Rule is minimized for a given p if the network eliminates that prime time program having the largest $\hat{p}$. Affiliation contracts currently call for about 30 hours of uncompensated time per month while prime time programming amounts to approximately 120 hours per month. Since only 15 hours of network programming must be eliminated, prime time hours having the largest $\hat{p}$ will lie to the right of OY. This conclusion may not hold in the case when prime time hours do not yield equal revenue.

[21] A network spokesman has estimated that a loss of

VII. The Group Ownership Rule and the Fourth Network

One of the perplexing aspects of behavior in the television broadcasting industry is the fact that very large profits are earned by affiliated television stations, yet the formation of a fourth commercial network has not taken place.[23] This has apparently resulted in very small profits for those stations unable to obtain an affiliation, and, in some cases, the failure to obtain an affiliation has forced stations out of business.

The formation of a network involves very substantial costs which are independent of the number of affiliates. In addition, there are relatively low costs of supplying an additional station with programming.[24] According to a well-known proposition in welfare economics, the undertaking of any indivisible activity is justified if a perfectly discriminating monopoly could profit from doing so. And that would seem to be the appropriate test for the introduction of a fourth network. But the results of our analysis are that networks will be unable to extract all of the extra benefits that they generate in markets in which the number of stations is smaller than the number of networks. And, further, while it would seem possible for networks to extract all quasi rents which they might generate in other markets, they apparently fail to do so either because they are not very efficient price discriminators, because of the ban on option time, or both. As a result, some of the benefits generated by the formation of a network accrue not to the network but to its affiliated stations. This fact may partly explain why a fourth network is not formed even though it may be economically justified.

One way to allow networks to appropriate all of the benefits which they generate, the outright ownership of affiliates by the network, is precluded by the Group Ownership Rule which restricts the number of stations that a network can own.

In addition to restricting the degree to which networks can appropriate the rents which they generate, some FCC regulations such as the maximum permitted sharing ratio and the Prime Time Rule simply reduce network profits without increasing the profits of affiliated stations. These regulations also reduce the likelihood that a fourth network will develop.[25]

It may seem paradoxical at first that permitting networks to receive larger profits can enhance the success of local stations. But the paradox can be resolved when one realizes that most, if not all, of the marginal television stations have no network affiliation, and that a network affiliation, even on terms which are very favorable to the network, may be better than no affiliation at all.

VIII. Conclusion

This paper has provided a model of the economic relationship between television

one-half hour of prime time could cost each network \$6–14 million per year in gross profits. See *Wall Street Journal*, Apr. 15, 1970, p. 2.

The FCC limitation of the proposed rule to the 50 largest markets is apparently based on the assumption the networks will continue to furnish their affiliates in other markets with network programming throughout prime time. Apparently some of the smaller stations fear that they will not be able to obtain network programming through prime time, *Wall Street Journal*, May 8, 1970, p. 4. See also the *Wall Street Journal* editorial, May 13, 1970, which expresses a similar fear.

[22] It is useful to point out that there exists a relationship between our analysis of the ban on option time, the prime time rule, and the ceiling to the sharing ratio, and the literature on the behavior of the regulated firm. See, for example, Harvey Averch and Leland Johnson. Unlike the Averch-Johnson model, where the constraint imposed on the regulated firm is in terms of the maximum rate of return that it is permitted to earn, in our model the profits of the networks are not directly regulated and the restrictions take other forms.

[23] There have been a number of abortive attempts to form such a network.

[24] The principal incremental cost is the AT&T interconnection charge.

[25] Another barrier to entry of a fourth network which has been alleged to be important is the structure of AT&T interconnection charges. On this point see, for example, Joel Dirlam and Alfred Kahn.

networks and their affiliates. The model was employed to examine the manner in which advertising revenues are shared. We were able to demonstrate that the use of uncompensated time and of option time both contribute to the ability of television networks to extract the quasi rents which would otherwise accrue to local stations as a result of their network affiliation. We also analyzed the impact of the ban on option time, the restriction on the maximum permissible sharing ratio, and the new Prime Time Access Rule. An important policy conclusion is that these rules in addition to the limitation on station ownership imposed by the Group Ownership Rule may be important barriers to the development of a fourth network.

REFERENCES

H. Averch and L. L. Johnson, "Behavior of the Firm Under Regulatory Constraint," *Amer. Econ. Rev.*, Dec. 1962, *52*, 1052–69.

H. M. Blake and J. A. Blum, "Network Television Rate Practices: A Case Study in the Failure of Social Control of Price Discrimination," *Yale Law J.*, July 1965, *74*, 1339–1401.

R. H. Coase, "The Federal Communications Commission," *J. Law Econ.*, Oct. 1959, *2*, 1–40.

J. B. Dirlam and A. E. Kahn, "The Merits of Reserving the Cost-Savings from Domestic Communications Satellites for Support of Educational Television," *Yale Law J.*, Jan. 1968, *77*, 494–519.

H. H. Goldin, "Economic and Regulatory Problems in the Broadcasting Field," *Land Econ.*, Aug. 1954, *30*, 223–33.

E. Greenberg, "Television Station Profitability and FCC Regulatory Policy," *J. Ind. Econ.*, July 1969, *17*, 210–38.

H. L. Jome, "Public Policy Toward Radio Broadcasting," *Land Econ.*, Apr. 1925, *1*, 198–214.

H. J. Levin, "Economic Effects of Broadcast Licensing," *J. Polit. Econ.*, Apr. 1964, *72*, 151–62.

J. L. Peterman, "The Structure of National Time Rates in the Television Broadcasting Industry," *J. Law Econ.*, Oct. 1965, *8*, 77–131.

Wall Street Journal, Apr. 13, 1962; June 1, 1962; Apr. 15, 1970; May 4, 8, 13, 1970.

U.S. Congress, House, Antitrust Subcommittee of the Committee on the Judiciary, Hearings: *Monopoly Problems in Regulated Industries, Part 2, Television*, 84th Congress, 2d sess., 1956.

On The Theory of Optimal Investment, Dividends, and Growth in the Firm

By CLEMENT G. KROUSE*

The consideration of time-optimal programs for corporate investments, dividends, and growth is a relatively new area of microeconomics. Although the specification of a maximand appropriate for such models has been well studied,[1] little systematic attention has been given to a model with suitable dynamic structure and, subsequently, to the intertemporal dependencies which relate the firm's value and earnings to decisions on financing and investments. It is the purpose of this paper to consider this second problem, particularly with regard to generalizing the conditions under which the theory of valuation can be used to determine the unlevered firm's optimal financing and investment decisions.

Previous formulations of dividend, investment, and growth programs for the firm have been restrictive in giving consideration only to one of two polar cases. One approach has identified dividend policy on a one-to-one basis with investment policy and thus confounded the two,[2] while the other has abstracted from the interdependencies between investment and financing decisions.[3] The difficulties associated with the first kind of abstraction have been fully outlined by Miller-Modigliani nearly a decade ago and need not be reviewed. The difficulties inherent in the second abstraction, although not considered by the literature in this context, are not less, for they set aside a fundamental proposition of modern capital analysis requiring that both (the firm's) production and exchange opportunities be jointly considered to yield a time optimal path for (shareholder) consumption claims.[4] To this end, this study explicitly and simultaneously treats the shareholder valuation functional and the firm's intertemporal earnings-possibility equation to make fully clear the conditions under which investment and financing opportunities, while distinct, interact to determine optima which involve unique and nonunique specifications of these two.

A further restrictive characteristic of prior corporate investment and financing studies is their a priori assumption of stationary-growth conditions. Particularly, they consider only the special equilibrium structure of corporate events where the economic state of the firm is identical in every period except by a constant scale factor, and all its decisions and exogenous variables (for example, discount rates) are required to be constant through time. Concomitantly, they limit the models to comparative dynamics, which is not useful for the analysis of the firm's decision rules or states as exogenous variables change over time, but only appropriate for the comparison of firms with different, time-constant values of these variables. While there

* Assistant professor of business economics, University of California, Los Angeles. This study was supported by the Research Program in Competition and Business Policy at University of California, Los Angeles.

[1] Notably by Jack Hirshleifer and Miller-Modigliani.

[2] See Gordon.

[3] See Miller-Modigliani.

[4] With a specified level of operations it is customary to show that the firm's financing decisions leave its total market value unchanged. More precisely, it is necessary to indicate, as herein, the conditions under which financing decisions do not affect the wealth position of initial shareholders. On this point, see Hirsheifer.

are differences in their developments, the models of James Walter, Myron Gordon, Merton Miller-Franco Modigliani, and John Lintner (1967) are typical of the stationary growth forms.[5]

The perspective of the firm taken herein is that of a dynamic, discrete-time activity. The firm is not a priori required to maintain constant growth rates in dividends or earnings, but is fully free to make financing and investment decisions consistent with optimizing initial shareholder wealth. Since our objective is to address specifically the optimal capital accumulation and financing problems of the corporation, production technology and product-factor demand functions are simply subsumed in the specification of an earnings-possibility function. Earnings are defined as income net of economic depreciation and, on the basis of this definition, a first-order difference equation is developed to relate their level at any point in time to prior levels and the size and financing source of capital budgets.

With regard to the objective function, the neoclassical theory of optimal capital accumulation is employed which implies using an objective maximizing the utility of a stream of consumption. Further, based on the proof of Hirshleifer that the maximization of the firm's present value is the only criterion fully in agreement with the external consumption stream objective, the frequently cited maximization of the average internal rate of return on invested equity, geometric rate of earnings growth, etc. are disregarded and the focus placed singularly on extremizing the discounted stream of dividends paid (over an infinite horizon) to the firm's "initial" shareholders.

[5] Particularly, the M-M "convenient case of constant growth rates" and the Lintner certainty model. Gordon and Walter, it should be noted, consider only the all retained-earnings financing assumption while the other studies also admit external equity financing.

Necessary and sufficient conditions for extremal earnings and dividend time sequences, along with a program of time-optimal retained earnings and external equity decisions, are developed based on this model. In important ways these extremal conditions extend to greater generality the commonly used marginal earnings efficiencies of the various financing sources and the frequently cited Miller-Modigliani proposition concerning the null effects of the financing mode on the firm's market value. Finally, the specific effects of technical progress and market discount rates on the optimal size of the capital sources in each period are developed.

I. The Model

While there is an increasing volume of theoretical literature examining the investment, financing, and growth of the firm, the failure to be fully explicit about specializing assumptions and to clearly set forth definitions of economic variables has unnecessarily led to misinterpretations of these analyses, particularly with regard to the generality of their conclusions. To this end this section sets out in detail the variables and important analytical components of the current model, in part so that the reader can be clear about the proper limits of its implications and in part to show the fundamental respects in which it generalizes prior studies.

First, define the usual variables of the firm.

$D(t)$ = *total* dollar amount of dividends paid at time t.

$D_T(t)$ = portion (in dollars) of time t dividends paid to shareholders of record at time T.

$k(t)$ = time t market discount *factor*, discounting back to time zero,

$$k(t) = \prod_{\tau=0}^{t} \{1/[1 + r(\tau)]\}$$

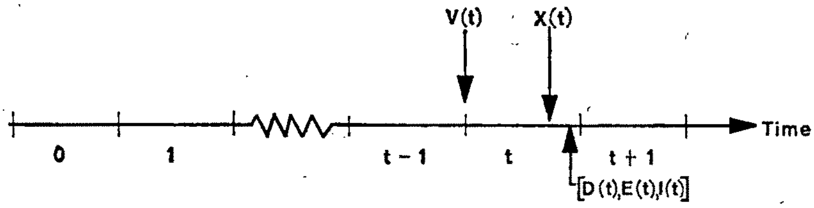


FIGURE 1. THE ECONOMIC PROCESS

where $r(\tau)$ is the one period opportunity rate available to the firm's shareholders.

$X(t)$ = time t net earnings of the firm (net of economic depreciation).[6]

$I(t)$ = transaction costless *internal* equity decision variable in period t—retained earnings for $0 \leq I(t) \leq X(t)$, with a preemptive ("rights") offering when $I(t) > X(t)$.

$E(t)$ = *external* equity decision in period t having associated transaction (brokerage) costs; expressed as the total dollar amount of the acquisition, $E(t) > 0$, or retirement, $E(t) < 0$, *including* transaction costs.

$\delta(E(t), t)$ = the ratio of the time t market price of an equity share to its external issue price: $\delta(t)$ for notational convenience. $\delta(t) < 1$ implies issuance of shares at a premium to market price, and conversely for $\delta(1) > 1$. Some expectations/information asymmetry among potential shareholders is held to account when $\delta(t) \neq 1$.

$V(t)$ = value of the corporation at the beginning of period t to shareholders of record at *that* time.

Taxes are everywhere neglected for convenience. The timing of the decisions and money flows is depicted in Figure 1. For a representative period t the sequence of occurrences is such that, after earnings are realized, all decisions are made at essentially period end. In keeping with convention, the value of the firm is established at the beginning of the period. Thus, new external equity is raised in every period *ex dividends* of that period.

Given these definitions, the value of the corporation at an arbitrary time, t = 0, to shareholders of record at that time, $V(0)$, can be represented by the functional

$$V(0) = \sum_{t=0}^{\infty} D_0(t)k(t) \tag{1}$$

The expression explicitly allows for the future issue of new external equity shares. This particular form, however, is of limited use since $D_0(t)$ is not an easily observable state or decision of the firm and it is accordingly convenient to transform the functional. Separating initial from later periods, rewrite equation (1) in the form.

$$V(0) = k(0)D_0(0) + k(0)\sum_{t=1}^{\infty} D_0(t)\hat{k}(t)$$

where $\hat{k}(t)$ now discounts back to the beginning of period 1, not 0.

Since new equity shares are issued at the end of period 0, and hence *ex dividends*, the entire initial-period payout is to the initial shareholders. But, only a pro rata portion of the total dividends paid in later periods will accrue to the shareholders of record at

[6] Precisely, the Hicks net income (No. 2) definition is utilized. See chapter 14.

time t=0. Accordingly, $V(0)$ becomes

$$V(0) = k(0)\left\{D(0) + \left[1 - \frac{\delta(0)E(0)}{V(1)}\right] \cdot \sum_{t=1}^{\infty} D_1(t)\hat{k}(t)\right\}$$

The term $E(0)$ is the total dollar amount of external equity acquired (including transaction costs), which may be at an average share price some premium ratio above market value, $\delta(0)<1$, below market value, $\delta(0)>1$, or at market value, $\delta(0)=1$. The final summation term in this expression is the dividend sequence accruing to shareholders of record in period 1 discounted to that time. This is exactly the value of the corporation in period 1, $V(1)$, so the full expression can be simplified to

$$V(0) = k(0)[D(0) - \delta(0)E(0)] + \sum_{t=1}^{\infty} D_1(t)k(t)$$

By recursively employing this procedure for all future periods and further substituting the identity $D(t) \equiv X(t)-I(t)$, current shareholder value over an infinite horizon is written[7]

$$(2) \quad V(0) = \sum_{t=0}^{\infty} [X(t)-I(t)-\delta(t)E(t)]k(t)$$

The contribution to $V(0)$, and hence to the share price to holders of record at t=0, owing to the firm's decisions in any period can be seen to be positive, zero, or negative as dividends are greater/less than the market value of external equity issued (retired), $X(t)-I(t) \gtrless \delta(t)E(t)$. Particularly, in periods when the per share dividend to *market* price ratio is less than the fractional growth in shares outstanding the contribution will be negative.

Although somewhat different in form, the equation (1) expression for $V(0)$ is fully consistent with the familiar Miller-Modigliani equation (3) "fundamental rule of valuation." Assumptions in the sequel concerning decreasing marginal returns to investment, rates of earnings-technical adjustment, and the time properties of the discount factor, $k(t)$ are such to assure the sum converges. Shareholder wealth as expressed by (2) is to be maximized by the choice of a time sequence of investment funding decisions, I and E, from which, insofar as a retained earnings sequence is determined by I, a given dividend strategy follows.

Turning to the firm's earnings time path, it is required to set forth the relationship between the firm's investment budget and mode of financing at every point in time and its resultant earnings *sequence*. Following directly from the definition of the firm's income as net of economic depreciation, the manner in which earnings "spread out" from given investments can be very simply specified by a change function ϕ and the first-order difference equation

$$(3) \quad \Delta X(t) = A(t)\phi[I(t), E(t)]$$

where t denotes time, and $\Delta X(t)$ is a forward difference:

$$(4) \quad \Delta X(t) = X(t+1) - X(t)$$

The term $A(t)$ represents an index of earnings-technical adjustment, which will be shortly described.

The earnings-possibilities function ϕ sets out the firm's one-period net earnings *change* arising from specific investments.[8]

[7] Since retained earnings has been considered the sole source of equity funds in a predominance of corporate growth models, direct discounting of the firm's *total* dividend stream has become generally (and wrongly) identified as the universally proper valuation criterion. As can be seen from the above developments, this is the case *only* for the firm not employing new external equity. In distinction, the form of equation (1) fully compensates the current market value for future dividends that must be diverted to additional equity.

[8] This relation is assembled from demand functions, production functions (or cost functions), and factor prices. It, of course, assumes the optimal allocation of

Particularly, in the case of equation (3) it sets forth the earnings change from t to t+1 owing to the size and financing source of the firm's capital budget at time t. As noted earlier, the new level of earnings, after the one-period increment, is kept intact (unentrenched) throughout all future time via "automatic" reinvestments defined as economic depreciation. Thus, for the firm with zero *net* investments an undegraded new income level is maintained *ad infinitum*. In keeping with conventional assumptions it is further supposed that $\phi(\mathrm{t})$ is continuous through second partial derivatives of $I(\mathrm{t})$ and $E(\mathrm{t})$, has the null property where $\phi[0, 0]=0$, and exhibits decreasing marginal returns to investment in some region.

In the sequel we write

$$(5)\qquad \Delta X(\mathrm{t}) = A(\mathrm{t})\phi(\mathrm{t})$$

where $\phi(\mathrm{t})$ is notational convenience for $\phi[I(\mathrm{t}), E(\mathrm{t})]$.

The index $A(\mathrm{t})$ of the earnings difference equation is a measure of the firm's time t rate of earnings adjustment arising from advancements in the physical technology of production and, more significantly perhaps, from changes in the structure of its factor and/or product markets.[9] Such adjustments are modeled to be autonomous and to occur at a relative rate, $a(\mathrm{t})$, unique to each period:

$$\Delta A(\mathrm{t}) = a(\mathrm{t})A(\mathrm{t})$$

It may be observed that the product $a(\mathrm{t})A(\mathrm{t})$ ties prior earnings-technical affect to the current rate of change. By suitable choice of measure it is further advantageous to require $A(0)=1$, thus

$$(6)\qquad A(\mathrm{t}) = \prod_{\tau=0}^{\mathrm{t}-1}[1 + a(\tau)]$$

Adjustments at time t will be referred to as positive, neutral, or negative depending on whether $a(\mathrm{t})$ is greater than, equal to, or less than zero, respectively.

As a final element of generality, it should be explicitly noted that "negative" retained earnings are permitted to accommodate disinvestment and that negative dividends are permitted to accommodate preemptive ("rights") offerings of shares to existing shareholders. The central feature of such rights issues is their essentially zero transactions costs (so that they are properly entered into the earnings-change function through the transactions-costless internal equity variable).

II. Model Solution

Collecting the equations and definitions introduced to this point, the requirement for a time optimal investment, dividend, and growth program for the firm can be stated as

$$(7\mathrm{a})\qquad \max_{I(\mathrm{t}),\,E(\mathrm{t})} \sum_{\mathrm{t}=0}^{\infty} [X(\mathrm{t})-I(\mathrm{t})-\delta(\mathrm{t})E(\mathrm{t})]k(\mathrm{t})$$

subject to:

$$(7\mathrm{b})\qquad X(\mathrm{t}+1) = X(\mathrm{t}) + A(\mathrm{t})\phi(\mathrm{t})$$

$$(7\mathrm{c})\qquad X(0) \text{ given}$$

$I(\mathrm{t})$ and $E(\mathrm{t})$ are the firm's internal and external equity decision variables and its net earnings level, $X(\mathrm{t})$, is termed a state variable.

Unlike most prior treatments, the completed model of equation (7) states in a clear, choice-theoretic structure the firm's investment and financing problem: the ob-

each marginal dollar of investment, both among existing and potential new markets. Note also that the earnings-possibilities function is written to depend not on a simple investment aggregate in the form $\phi[I(\mathrm{t})+E(\mathrm{t})]$, but more generally by the $\phi[I(\mathrm{t}), E(\mathrm{t})]$ functional relationship. In this way transaction costs associated with acquiring external equity and any relationship between retained earnings and the cost of external funds can be fully reflected.

[9] Changes in product-factor market structure are modeled to affect the firm's earnings only by affecting its investment possibilities—in the sequel it will be seen that the optimal investment rules involve $A(\mathrm{t})$. Product factor-market changes which do not induce investments thus do not give rise to earnings adjustments.

jects of choice are specifically identified; a preference function ordering the choice-objects and resultant economic states is explicit; and, the time-process equation linking the firm's earnings state and decisions is set forth in detail. From this analytical form several distinguishing elements, vis-à-vis prior formulations, can be brought into clearer focus: internal and external equity financing modes are both available; financing decisions are not required to be constant over time, but can be uniquely chosen in each period; the usual constant "internal rate of return" is not imposed on the firm's investment possibilities; neither earnings nor dividend time sequences are a priori required to have any specific time structure; the firm's optimal investment and financing programs are to be jointly and consistently derived (the *total* investment or capital budget results as the net sum of funds from the two capital sources); and, finally, a measure of technical progress in the firm's earnings possibilities is explicitly incorporated, issuance of external equity at above or below market value is admitted, and the market discount rate is not assumed constant but as having a fully flexible time structure.

Looking to the content of equations (7), the firm's problem is to find the *sequences* I^* and E^* that maximize the value to current shareholders. Derivation of such an optimal program can proceed first by adjoining the earnings difference equations to $V(0)$ with a multiplier sequence λ.

$$\overline{V}(0)=\sum_{t=0}^{\infty}\{[X(t)-I(t)-\delta(t)E(t)]k(t) + \lambda(t+1)[X(t)+A(t)\phi(t)-X(t+1)]\} \tag{8}$$

For notational convenience, further define the Hamiltonian sequence over $t=0, 1, 2, \ldots$

$$H(t) \equiv [X(t) - I(t) - \delta(t)E(t)]k(t) + \lambda(t+1)[X(t) + A(t)\phi(t)] \tag{9}$$

The Hamiltonian at any time t is seen to be a sum of that period's discounted change in initial valuation, $V(0)$, and a product term of the auxiliary variable $\lambda(t+1)$, and earnings at time $t-1$. Thus the auxiliary variable has the interpretation of an intertemporally imputed value (shadow price) of an additional dollar of earnings (income) measured in terms of the present value of initial shares.

Finally, change the indices of summation for the $\lambda(t+1)X(t+1)$ term in equation (8) to obtain the notationally simpler expression for $\overline{V}(0)$.

$$\overline{V}(0) = H(0) + \sum_{t=1}^{\infty}[H(t) - \lambda(t)X(t)] \tag{10}$$

In this form solution of the problem can proceed directly by the discrete maximum principle, see Samuel Katz. Choose the convenient multiplier sequence

$$\lambda(t) = \frac{\partial H(t)}{\partial X(t)} = \lambda(t+1) + k(t) \tag{11}$$

which importantly determines the auxiliary variable only as a function of the market discount rates and not as being dependent on any of the firm's decisions or states. Moreover, with the "free end" conditions $\lambda(t)$ becomes zero in the limit as t becomes infinite. Since by (11) $\Delta\lambda(t) = -k(t)$, then $\lambda(0)$ is given as

$$\lambda(0) = \sum_{t=0}^{\infty} k(t) \tag{12}$$

If the market discount rates are always positive, then the $k(t)$ discount factor becomes zero as $t\rightarrow\infty$, and the above summation converges to a finite value—which value simply becomes one over the discount rate itself when that rate is, as in all prior models, considered constant over time. Finally equations (11) and (12) can be combined recursively to yield

$$\lambda(t+1) = \sum_{\tau=t+1}^{\infty} k(\tau) \tag{13}$$

which is a forward sum of discount factors, discounting to $t=0$.

By the maximum principle it is necessary for an optimal financing program that the firm choose $I(t)$ and $E(t)$ to

$$(14) \quad \max H(t) \qquad \text{for all } t = 0, 1, \ldots$$

Important economic implications of this condition become clear if equations (9) and (13) are utilized to write the Hamiltonian as

$$(15) \quad \begin{aligned} H(t) &= [X(t) - I(t) - \delta(t)E(t)]k(t) \\ &\quad + [X(t+1)] \sum_{\tau=t+1}^{\infty} k(\tau) \end{aligned}$$

Thus, it is seen that the maximum principle requires the firm *at each period in time* to maximize a properly weighted sum of a *direct* contribution to market value, $[X(t) - I(t) - \delta(t)E(t)]$, and an *indirect* contribution by the process of augmenting its earnings state, $X(t+1)$; the weighting factors are, respectively, $k(t)$, which represents the present value of a marginal change in the firm's time t dividend and external equity acquisition decisions, and the sum of $k(\tau)$, $\tau = t+1, \ldots$, which is the present value of a marginal change in the firm's net income stream. Indeed, this is *exactly* what is meant when it is said the firm has a static investment and financing problem. To the degree that the firm pays dividends and does not acquire external capital in any period, it operates to increase $V(0)$ directly, but simultaneously reduce the potential indirect contribution to value via future earnings. In a reverse manner, the firm might alternatively choose to reduce its direct, but increase its indirect, contribution to value. A "proper balance" in the two effects is, of course, necessary for optimality. The maximum principle yields a precise statement of the firm's multiperiod investment and financing problem in a derived one-period equivalent by, most importantly, supplying the proper weighting values to these direct (immediate) and indirect (future) effects.

Stationarity conditions for the maximization problem given by (15) require at each period of time that

$$(16a) \quad -k(t) + \lambda(t+1)A(t)\frac{\partial \phi(t)}{\partial I(t)} = 0$$

$$(16b) \quad -k(t)M(t) + \lambda(t+1)A(t)\frac{\partial \phi(t)}{\partial E(t)} = 0$$

where $M(t)$ is the marginal external issue market value (rate), $\partial[\delta(t)E(t)]/\partial E(t)$. The solution to these equations are necessary and sufficient for a unique maximum to the problem if $H(t)$ is strictly concave. This property is generally assumed since ϕ is typically subject to diminishing marginal returns, and as $E(t)$ rises the marginal external issue market value will typically increase (meaning the issue needs to be "sweetened" at an increasing rate with its size).[10]

With regard to structure, a key property of the solution equations is their "decomposition" of the firm's decisions over time. In properly imputing a value to future earnings by the intertemporal shadow value $\lambda(t+1)$, each period's decisions are developed such that they can be optimally and "time-independently" undertaken. In an important way this specific feature of the solution technique distinguishes the current analysis from earlier studies which decouple the firm's time sequence of decisions more simply and restrictively by requiring that economic conditions in each

[10] Let $\phi_{II}(t)$ and $\phi_{EE}(t)$ be the second partial derivates of $\phi(t)$ with respect to $I(t)$ and $E(t)$, and let $D_\phi(t)$ be the Hessian determinant of $\phi(t)$ by those decision variables. Sufficiency conditions particularly require (16) to hold, $\phi_{II}(t)<0$, $\phi_{EE}(t)<k(t)(\partial M(t)/\partial E(t))$, and $\partial M(t)/\partial E(t) > D_\phi(t)/k(t)\phi_{II}(t)$. With diminishing marginal transactions costs for external equity $D_\phi(t)<0$ and the last condition on the minimal rate at which the marginal external issue market value increases must be assumed.

period be identical with those of every other.

Written more compactly, a maximum for shareholder value requires the firm at each period of time to select $I(\mathrm{t})$ and $E(\mathrm{t})$ financing to solve

$$\text{(17a)} \qquad \frac{\partial\phi(\mathrm{t})}{\partial I(\mathrm{t})} = Z(\mathrm{t})^{-1}$$

$$\text{(17b)} \qquad \frac{\partial\phi(\mathrm{t})}{\partial E(\mathrm{t})} = M(\mathrm{t})Z(\mathrm{t})^{-1}$$

where for notational convenience in the sequel $Z(\mathrm{t})$ is introduced, being defined as

$$\text{(18)} \qquad Z(\mathrm{t}) \equiv A(\mathrm{t}) \sum_{\tau=\mathrm{t}+1}^{\infty} \frac{k(\tau)}{k(\mathrm{t})}$$

$Z(\mathrm{t})$ will be termed the *adjusted* sum of *forward* discount factors.[11] The decision rules given by (17) are thus not myopic since they involve the parameters of future periods, and static theories of the firm's investment and financing problem are invalid unless they employ these proper intertemporal prices.

The following two sections are concerned with the content of equations (17) as optimal investment decision rules for the firm and for the implications they have with regard to the uniqueness of the firm's financing "package."

1. *Optimal Investment Decision Rules*

Notwithstanding Lintner's "current earnings yield" decision rule as a cutoff rate for investment owing to his particular definition of earnings, the familiar "valid and appropriate" criterion for determining the optimal size capital budget has the firm investing until its marginal internal rate of return equals the *constant* external market discount rate.[12] In extension to this traditional rule, the prescriptions of equation (17) are more generally appropriate, for they permit a freely given time structure for the firm's discount rate and further accommodate situations with nonneutral technical change.

It might be replied that current theories do permit a time-varying discount rate structure. For, the most commonly employed model of the firm's intertemporal investment and financing problem develops, by construction, an "equivalent composite" market discount rate, r_c, as "the" root of the equation

$$\sum_{\mathrm{t}=0}^{\infty} \frac{1}{(1+r_c)^{\mathrm{t}+1}} = \sum_{\mathrm{t}=0}^{\infty} \frac{1}{\prod_{\tau}^{\mathrm{t}} [1+r(\tau)]}$$

When this technique is used, however, the effect of the short rates, $r(\tau)$ for *all* $\tau=0, 1, 2, \ldots$, are embodied in r_c and particularly such that the weight given early periods in the horizon is greater. Thus, when the compositing scheme is used to simplify the problem the "invest-down-to-the-constant-market-rate" decision rule for an equity cutoff is affected by the market rates of all prior and future periods. In contrast, the decision rules developed here —equations (17)—properly show that only *future values* of the discount rates are relevant as a basis for establishing the proper marginal return from equity. In a reverse fashion, the form of $A(\mathrm{t})$, see equation (6), as it enters these decision rules means that only *prior* earnings-technical adjustment affects the optimal equity "cutoff" with no future values, or expectations of them, being relevant.

[11] Note:

$$\sum_{\tau=\mathrm{t}+1}^{\infty} [k(\tau)/k(\mathrm{t})] = \frac{1}{1+r(\mathrm{t}+1)} + \frac{1}{[1+r(\mathrm{t}+1)][1+r(\mathrm{t}+2)]} + \ldots$$

where $r(\mathrm{t}+1)$ is the market discount *rate* of period t+1. Thus, this is properly a *forward* sum of discount factors, each element discounting back to the end of period t.

[12] See Miller for a defense of this rule, but also a warning that it is based on assumptions which are "terribly treacherous."

Turning directly to an inspection of equations (17), the decision rules for an optimal investment program require: 1) that the marginal return to zero-transactions-cost internal equity, $\partial\phi(t)/\partial I(t)$, equal the inverse of the adjusted sum of forward discount factors and 2) that the marginal return to external equity issues, or any equity having transactions costs associated, $\partial\phi(t)/\partial E(t)$, always equal the *prevailing* marginal external issue market value times that same inverse of the adjusted sum of forward discount factors. Future expectations of an equity issuance premium or discount are immaterial. Optimal $I^*(t)$ and $E^*(t)$ follow from the simultaneous solution of equations (17).

If $\delta(t)\equiv 1$, implying that shares are issued and subscribed at market value, and marginal transactions cost are positive for any external issue, then external sources of funding will not optimally be employed, $E(t)\equiv 0$. External financing will only be used in the circumstances where a range, over $E(t)$, exists such that the marginal external issue market value is less than the marginal contribution of an external issue to investable funds. (The issuance of stock at a premium to market value results in a *gain* to existing shareholders and thus affects the tradeoff between internal and external generated equity.) Letting J be the investable portion of externally acquired funds (net of transactions costs) then $\partial\phi/\partial E=\partial\phi/\partial J\cdot\partial J/\partial E$ and $\partial\phi/\partial J\equiv\partial\phi/\partial I$: thus, at the optimum $M(t)=\partial J(t)/\partial E(t)$, the equality of the external issue marginal market value and marginal contribution to investable funds, is necessary.

These new rules are consistent as extensions to the standard (comparative dynamic) treatments of the firm's optimal investment decision. Indeed, it is of interest to see the manner in which they specialize under the several restrictions of prior studies. Particularly, suppose the commonly employed circumstance of constant market discount rates, neutral earnings-technical adjustment, and all new shares issued precisely at market price. From equation (18) it follows that $Z(t)=1/\bar{r}$, where $\bar{r}$ is the constant discount rate. Further not distinguishing acquired equity to permit associated transactions cost implies $I(t)+E(t)\equiv B(t)$, where $B(t)$ is an investment aggregate. Thus, the earnings-change function simply becomes $\phi(t)=\phi[B(t)]$, and the "minimum" cost financing can be subsequently chosen. With these restrictions, equations (17) degenerate to the familiar rule that $\partial\phi(t)/\partial B(t)=\bar{r}$ and, as common, the rule simply requires that the firm optimally invest "down" to *the* external market rate.

2. *Dividend Relevancy and Conditions for a Unique Financing Program*

Turning now to the dividend relevancy question popular in both the theoretical and empirical literature, the form of the present model makes it possible to adduce more precisely the conditions under which a unique dividend and financing program exists. As a precursor to complete conditions it is first convenient to sketch the essence of the Miller-Modigliani proof of the immateriality of the equity financing package. (A representative time t is used, but, of course, t could indicate any and every period.)

Specifically, as M-M, we wish to derive the conditions for changes in $I(t)$ and $E(t)$ which lead to a null effect on $V(0)$. The differential of $\bar{V}(0)$ according to $I(t)$ and $E(t)$ can be simply formed from equation (8) and set to zero yielding

$$\text{(19)}\quad \left[Z(t)\frac{\partial\phi(t)}{\partial I(t)}-1\right]dI(t)=\left[Z(t)\frac{\partial\phi(t)}{\partial E(t)}-M(t)\right]dE(t)$$

If we suppose that $\delta(t)\equiv 1$, $M(t)\equiv 1$ and $\delta\phi(t)/\partial I(t)\equiv\partial\phi(t)/\partial E(t)$, which is surely

guaranteed by the M-M assumptions of "perfect" capital markets, then the conclusions are subsequently clear: so long as the firm substitutes retained earnings for external equity on a one-to-one basis, or conversely, the firm's initial value will not change.[13] And, since dividends are the complement of retained earnings, the specific payout at any point in time is immaterial in establishing the value of the corporation.

The structure of the present model and its solution equations offer a convenient means to extend and clarify the M-M conditions. Interpreting the "rational behavior" postulate of M-M, p. 412, particularly to mean that the firm not only desires but positively acts to achieve maximum shareholder wealth, it is seen that the conditions for a unique, optimal financing program to exist are indentical to those under which a unique solution exists for equations (17). This existence question follows directly from the implicit function theorem: under circumstances in which the Jacobian of equations (19) is nonsingular in the neighborhood about the candidate optima, there exists a solution for $I^*(t)$ and $E^*(t)$ solely in terms of model parameters. The relevant Jacobian is merely the Hessian of $H(t)$ by the two decision variables. As a result, the role of both the firm's investment possibilities and transaction costs in specifying $\phi(t)$ and the properties of the external issue discount/premium function, $\delta(t)$, are important to the question of a unique financing package. Using subscripts to indicate partial differentiation and dropping time arguments, conditions for nonunique financing are $Z^{-1}M_E = (\phi_{II}\phi_{EE} - \phi_{IE}^2)/\phi_{II}$ at the optima. One idealized case in which these conditions hold is that of perfect markets—rationality, identical investor expectations ($\delta \equiv 1$), and no transactions costs ($\phi_E \equiv \phi_I$). These are the assumptions employed by M-M and can be simply seen to guarantee the singularity of the Hessian at the optima. It, however, should be noted that their assumptions are a sufficient, but not necessary, condition for this singularity and thus for the "irrelevancy" of dividends.

IV. Conclusions

Although the economic implications of the foregoing model and its development have been continuously noted, several stand out for special emphasis. First, and at the center of the entire analysis, is the precise statement of the firm's financing problem given by the requirement to maximize the Hamiltonian. The importance of this condition arises not from the fact that "direct" and "indirect" effects must enter into the investment and financing dedecision, and that these must be compromised, but, rather, it stems from the explicit weights given these two effects. A second key implication of the analysis is that dividend policy does generally affect the maximizing firm's value. Only under those specializing assumptions which give singularity to the relevant second-order Hessian determinant are dividends, or the financing package, irrelevant. Finally, the standard "invest-down-to-the-external-discount-rate" theorem has been extended to further specify the particular investment and financing program which is optimal, to accommodate an index of technical progress, to permit external equity to be acquired at a premium or discount with respect to its market value, and to reflect nonconstant external discount rate structures.

These and the several other conclusions of the present study importantly result from a generalization of prior models and not, as common, from considering further restrictive conditions. Thus, perhaps more

[13] Of course, if the firm is maximizing $V(0)$ under these specific assumptions then $\partial\phi/\partial I = \partial\phi/\partial E = Z^{-1}$ and the rate of substitution is ambiguous. We consider this case in the sequel. Also, note that if $\delta \neq 1$ then the rate of substitution will not be one-to-one.

useful than any of the specific implications derived, the current model lays a firm groundwork for further study. For, by giving a more analytic and flexible structure to the firm's decision problem a new basis is laid for extending the set and treatment of involved variables. An obvious first step in this direction would be to include debt financing as an argument in ϕ and then permit functional dependence between the firm's decisions (and concomitant changes in leverage ratio) and its appropriate external discount rate. Also the introduction of random elements for the development of optimal adaptive decision strategies might be highly fruitful.

REFERENCES

M. Gordon, "Dividend, Earnings, and Stock Prices," *Rev. Econ. Statist.*, May 1959, *41*, 99–105.

J. Hicks, *Value and Capital*, 2d ed., Oxford 1945.

J. Hirshleifer, "On the Theory of Optimal Investment Decision," *J. Polit. Econ.*, Aug. 1958, *64*, 329–52.

S. Katz, "A Discrete Version of Pontryagin's Maximum Principle," *J. Electronics Control*, May 1962, *13*, 46–57.

J. Lintner, "The Cost of Capital and Optimal Financing of Corporate Growth," *J. Finance*, May 1963, *28*, 292–310.

M. Miller, "Comments: Capital and Optimal Financing of Corporate Growth," *J. Finance*, May 1963, *23*, 314–17.

——— **and F. Modigliani**, "Dividend Policy, Growth and the Valuation of Shares," *J. Bus., Univ. Chicago*, Oct. 1961, *34*, 411–33.

J. Walter, "Dividend Policies and Common Stock Prices," *J. Finance*, Mar. 1956, *11*, 29–41.

Private Demands for Public Goods

By THEODORE C. BERGSTROM AND ROBERT P. GOODMAN*

Goods and services which are publicly supplied generally share two features. Costs of these commodities are divided among members of the community. Decisions about quantities to be supplied are made collectively. Somehow, from a community consisting of many individuals with different tastes, different amounts of wealth, and conflicting interests, a single decision must emerge. Quantities must be determined and costs divided in some way.

In this study we develop a method for estimating demand functions of individuals for municipal public services. Demand will depend on the traditional economic variables, price and income, as well as on certain demographic characteristics of the individual and the city in which he lives. Knowledge of individual demand functions would be useful for many purposes. Such uses include:

1) Prediction of the outcomes of alternative political decision methods and tax structures in a particular city.

2) Computation of tax structures and expenditure levels which satisfy certain preference-based normative criteria such as the Lindahl equilibrium. This would enable one to make statements about whether "too many" or "too few" public goods are actually provided and whether tax burdens are divided "equitably" relative to the norm employed.

3) Investigation of whether there are "scale economies" to city size in the production of public services.

4) Prediction of the effects of projected changes in values of economic and demographic variables on quantities of public goods to be supplied.

In the case of privately purchased commodities, it is in principle possible to observe directly the choices made by individuals belonging to a particular demographic group under alternative price and income situations. This is not in general the case with publicly supplied commodities.[1] One is, however, able to observe the choices made by many distinct municipalities. Employing rather simple assumptions on the nature of the political process in each municipality, this information can be used to make inferences about the responses of individual demands for municipal services to price, income and other variables.

As an illustration of how this may be done, consider the following assumptions:

I. Units of measurement for a given municipally supplied commodity can be chosen in such a way that each municipality j is able to supply the commodity at constant unit cost q_j.[2]

* Associate professor, department of economics, Washington University, St. Louis and economist, department of economics and human development, State of Maryland, respectively. This research was done at Washington University and was in part supported by NSF Research Grant GS-3070. A more detailed presentation of the results is available on request. We are grateful for assistance and helpful comments from James Barr, Banarsi Dhawan, John Legler, and Richard Meadows.

[1] This problem and related issues are well discussed by James Buchanan and Gordon Tullock.

[2] This will be possible even if communities produce public goods using some local inputs whose prices may differ from place to place, so long as a) all municipalities have identical, homothetic production functions and b) all municipalities face horizontal supply curves for inputs. To see this, simply choose output units so that production takes place at constant returns to scale and observe that for any community, a change in output so measured results in a proportional change in total costs. Of course unit costs may differ from place to place, but in each place unit costs are constant with respect to output.

II. For each consumer i there is a tax share τ_i such that i must pay the fraction τ_i of the total cost of municipal expenditures in his community. Consumer i's tax share may depend on his wealth, his income or other individual characteristics, but it does not vary with the size of municipal expenditures nor with the way in which he expresses his desires for municipal services.

III. Any consumer i living in municipality j is aware of his "tax price," $\tau_i q_j$ and is able to determine the quantity of the municipal commodity which he would choose for the community given that he must pay the fraction τ_i of its total cost. To do so he needs only to maximize his own preferences subject to a linear budget constraint where his tax price for the municipal good is $\tau_i q_j$.

IV. In each municipality, the quantity supplied of the municipal commodity is equal to the median of the quantities demanded by its citizens.

V. In each municipality the median of the quantities demanded is the quantity demanded by the citizen with the median income for that municipality.

Observe that if Assumptions IV and V are true, the quantity of municipal commodities chosen by any community is the amount which is desired by the consumer with the median income for that community. Our assumptions about the constancy of τ_i and q_j, enable us to treat the expenditures of any municipality as an observation on the demand curve of a consumer with the median income for this municipality, where the price he pays for public goods is proportional to his tax share.[3] Assumptions I–IV are proposed and studied by Howard Bowen and independently by James Barr and Otto Davis. Assumption IV will hold if, for example, there is a majority voting system in which the quantity selected is such that half the population wants more and half the population wants less than that quantity.

DEFINITION: *An allocation satisfying I–IV will be called a Bowen equilibrium.*

Circumstances under which V is justified will be discussed below. It will also be demonstrated that our estimation procedure applies under somewhat weaker assumptions than V.

I. Estimating Procedure

Our sample consists of 826 municipalities with 1960 populations between 10,000 and 150,000, located in 10 states.[4] Multiple regression is used to fit a function of the form:

$$\log E = c + \alpha \log n + \delta \log \hat{\tau} + \epsilon \log \hat{Y} + \sum_{i=1}^{k} \beta_i X_i$$

The symbols are defined as follows:

E = The expenditures of a municipality on a specified category of municipal service. In particular, separate regressions were run for the following definitions of E.

(a) E = Police expenditures.

[3] A similar approach to the estimation of individual demands for public goods is suggested by Robin Barlow, by James Barr and Otto Davis, and by Thomas Borcherding and Robert Deacon.

[4] The states included and the number of observations from each are shown in Table 1. Our choice of states was determined in part by availability of data and in part by a desire to include states which had a large number of municipalities. The main limitation of data is the availability of figures on equalized assessed valuation by municipality. (Connecticut was excluded because of a lack of other demographic data.) There are several "small" states for which comparable data is available but which were not studied because of limitations of time. Results are reported for all states which were studied. It was believed that cities with populations exceeding 150,000, because of their relative small number and because of the more detailed information available concerning them, would be more appropriately treated in a separate national cross-section study. Places with populations below 10,000 were excluded because of a lack of comparable data.

(b) E = Parks and recreation expenditures.

(c) E = Total municipal expenditures excluding education and welfare.

n = The number of households in a municipality.

$\hat{\tau}$ = The tax share of the citizen with the median income for a municipality.

$\hat{Y}$ = Median income in a municipality.

The X_i's consist of the values of several descriptive social and economic variables for a municipality. These variables will be discussed below. As argued previously, Assumptions I–V would permit us to use the expenditure of any municipality as an observation of the quantity demanded by one of its citizens whose income is $\hat{Y}$ and whose tax share is $\hat{\tau}$.

If all municipally supplied commodities were pure public goods as defined by Paul Samuelson, the use of a public commodity by any consumer would not reduce its usefulness to others. If this were the case, the size of a city would affect the quantity of municipal commodities only indirectly through the tax share. In larger cities, costs would be divided among more citizens and tax shares would tend to be smaller, hence individuals would tend to vote for more public commodities.

If there is crowding of municipal services, population size may also have a direct effect on individual demands. We suggest here a simple model of crowding which lends a natural interpretation to our population coefficients. Suppose that the usefulness of a public facility to any individual is determined by a function of the form $Z^* = n^{-\gamma}Z$ where n is the number of people sharing the good and Z is the quantity of the public good. Let the utility function of any individual, i, be of the form $u_i(X_i, Z^*)$ where X_i is the quantity of private goods which he consumes. If $\gamma = 0$, the public good is a Samuelsonian pure public good. If $\gamma = 1$, the individual's preferences are as if he received and enjoyed only the fraction $1/n$ of the total amount of the public good. Suppose the unit cost of the private good is p, the unit cost of the public good is q, and the tax share of consumer i is τ_i. Then his demand function is found by maximizing $u_i(X_i, Z^*)$ subject to

$$X_i + \tau_i qZ \leqq Y_i$$

or equivalently subject to

$$X_i + \tau_i qn^\gamma Z^* \leqq Y_i$$

Thus, determination of his demand function for Z^* is formally equivalent to finding an ordinary demand function where the price is $\tau_i qn^\gamma$. Let us suppose that there are constant income and price elasticities δ and ϵ for the commodity Z^*. Then the demand function for Z^* is of the form,

$$c[\tau_i qn^\gamma]^\delta Y_i^\epsilon$$

The quantity of Z demanded is n^γ times the quantity of Z^* demanded. Hence, his demand for Z is

$$n^\gamma c[\tau_i qn^\gamma]^\delta Y_i^\epsilon = cq^\delta \tau_i^\delta Y_i^\epsilon n^{\gamma(1+\delta)}$$

Thus, the coefficient, α, for the elasticity of demand with respect to population could be interpreted to be $\gamma(1+\delta)$.[5]

If γ were nearly zero, there would be substantial economies to larger city size since in larger cities, more consumers could share in the costs of municipal commodities with only minor crowding effects. Where γ is about one, the gains from sharing the cost of public commodities among persons are approximately balanced by the disutility of sharing the facility among more persons.

The values of the auxiliary variables

[5] A very similar analysis of crowding is presented by Thomas Borcherding and Robert Deacon. Notice that if $\gamma = 1$, the demand equation reduces (in *log* form) to $log\ (E/n) = log\ c + \delta\ log\ (q\tau_i n) + \epsilon\ log\ Y_i$. In this case we could consider demands for *per capita* quantities as a function of tax price *multiplied by population* and of income.

$X_1, \ldots, X_k$, may have two distinct kinds of effects on the expenditure decision of a municipality. Some variables appear to describe characteristics of a city which have fairly uniform effects on the demands of all its citizens for public services. Other variables seem to enumerate a portion of the population which may either have different tastes from the remainder of the population or pay a different tax share than persons with similar income. Examples of the former type are density and employment-residential ratio. Examples of the latter type are percent owner occupied, and percent of population older than 65. In practice, it is difficult to make this distinction in an unambiguous way. (How would one classify population change or percent nonwhite?) The first type of effect presents no difficulties for our method of analysis. The second type presents problems which will be discussed below. Since we have employed only a few of a large complex of interrelated social variables, it is difficult to attribute a clear-cut causal interpretation to the coefficients of these variables. At best, we hope that the variables which we include are adequate to eliminate substantial distortions of our estimates of price, income, and population elasticities.

II. Measurement of Variables

With the exception of the tax share variable, all of our data are available from standard sources. Roughly speaking, the income and demographic data are available from the 1960 census of population, while the expenditure data are from the 1962 census of governments.

We measure the quantity of municipal services by dollars of expenditure, median income in dollars, and the tax price of the median income consumer by his tax share. For this procedure to be literally consistent with our interpretation of the regression coefficients of income and tax share as price and income elasticities, we must assume that prices of private goods and unit costs of public goods are the same in each community observed. It is probably not unreasonable to assume that most of the supplies purchased by municipalities are purchased competitively in the same national market. But some of the inputs of municipal operations will be local labor and local land. Prices of these inputs may differ between communities. Local prices of *private* goods will also depend on local factor prices. Ideally we would like to deflate income by a local price index for all goods and deflate public expenditures by a local price index for public goods. We also would wish to use as our tax price variable the tax share *multiplied by the ratio of local public goods prices to local private goods prices.* Such local price indices are not, however, currently available.[6] It might not be unreasonable to suppose that the ratio of prices of public to private goods differs little between cities. In this case error enters our equation only because of the lack of a local price deflator for our income and expenditure figures. If the income elasticity of demand is close to unity, this will not seriously bias our estimates.[7]

[6] It would be possible in principle to construct such indices, using consumer budget studies, technical information about input requirements, and data on local factor prices.

[7] If q and p represent local prices for public and private goods, respectively, then the equation we would like to estimate is:

$$\log \frac{E}{q} = c + \alpha \log n + \delta \log \left(\hat{\tau} \frac{q}{p}\right) + \epsilon \log \left(\frac{\hat{Y}}{p}\right) + \Sigma \beta_i X_i$$

Rearranging terms we have:

$$\log E = c + \alpha \log n + \delta \log \hat{\tau} + \epsilon \log \hat{Y} + \Sigma \beta_i X_i + (1 + \delta) \log q - (\delta + \epsilon) \log p$$

This expression differs from the equation we estimated only by the final two terms. If q and p were uniform across cities, the omitted terms would affect only our intercept and not our estimates of elasticities. If the ratio of p to q were the same in every community then

Measurement of the tax share variable presents some conceptual and practical difficulties. We should like our demand equation to represent the response of individuals to the prices and incomes which they *perceive*. If one were to ask several individuals what their tax shares are, we suspect that few would be able to answer the question sensibly without more reflection than usually takes place before voting. Nevertheless, they may have knowledge which is for our purposes equivalent. For example, if a citizen knows his municipal tax bill and believes that his taxes will change in proportion to municipal expenditures, he will know the cost to him of a given percentage change in municipal expenditures. Moreover, it is often the case in municipal elections that an expenditure proposal is combined with a proposed change in property tax rates. A citizen who has some knowledge of the assessed value of his property can then determine the cost to him of the proposal.

It seems likely that although individuals have some notion of the cost of a proposal to them, their perceptions of tax shares may be quite imprecise. We formalize this idea by assuming that the perceived tax shares of individuals are independently distributed random variables with expected values equal to the actual tax shares. It turns out that in large populations the effects of independent errors of perception tend to cancel each other and that statistical distortions in our estimates from this cause are likely to be negligible. A formal statement of this proposition and its proof is offered in the Appendix.

In most cities sampled, more than half of locally generated revenue comes from the property tax. Property tax revenue is raised largely from taxes on real property (buildings and land) with the remainder derived from personal property taxes. To estimate the share of the tax on real property which is paid by a citizen with the median income for his community, we assume that in any municipality, the citizen with the median income lives in and owns the house with median value and that his house constitutes his entire holdings of real property. For each municipality, median house value is found in the 1960 U.S. Census of Housing. Tax rates and ratios of assessed to market value are determined for each community from information compiled by the states. The tax bill on the house of median value is then computed. This figure is divided by total property tax revenue for the municipality to produce an estimate of the share of real property taxes paid by the consumer with the median income.[8] It is assumed that the consumer with the median income pays the same share of other municipal revenues as he does of the real property tax. This is purely an assumption of conve-

the last two terms of the above equation would reduce to $(1-\epsilon)$ $\log p$ (plus a constant). If ϵ is near unity, then this term becomes insignificant.

[8] Those differences in median tax share from place to place which are not attributable to differences in population are largely due to differences in the amount of commercial and industrial property. Implicit in our procedure for estimating the tax share of the median consumer is the assumption that the burden of property taxes paid by commercial and industrial establishments is borne by persons living outside the municipality (or at least by an electorally insignificant portion of the cities' population). The assumption that individuals believe that their tax shares will not be altered if expenditures increase implies that they assume that the percentage of total revenue which can be extracted from commercial and industrial property will not change with changes in total revenue and expenditures. This may be somewhat unrealistic. A more thorough investigation might consider the question of the ability of a municipality to extract locational rents in an economy where the decision of a firm to locate in the municipality depends on levels of taxation and the quantity and composition of municipal taxation. Very little of the revenue for the services we study comes from other governments. The only exception is local highway expenditures for which about 15 to 25 percent of revenue comes from state collected funds. Even here, the transfers seem generally to be of the lump sum type, thus producing only a small income effect and no substitution effect.

nience which should be modified wherever better information is available.[9]

Although independent errors of perception by individuals are unlikely to have substantial effects on the parameter estimates, we are left with the problem that our own estimates of tax prices may differ substantially from the "true" tax prices. There are several reasons for this. One problem is that the accuracy of our estimates depends on the accuracy of the assessment to market value ratios computed by state equalization boards. These may be subject to substantial error. The assumption that the tax share of the consumer with median income is equal to his share of the property tax introduces further possibilities for error, as does the assumption that total unit costs of public goods are the same in all municipalities. Finally, there may be systematic nonindependent errors of perception in a community depending on the information made available to its citizens by the municipal authorities. If there are errors of measurement in the calculation of tax shares, the regression coefficients will be biased estimators and tend to underestimate the absolute value of the price elasticity. We therefore will compute maximum likelihood estimates of the price elasticity under some alternative assumptions about the measurement error.

III. Legitimacy of the Estimating Procedure

Suppose that the tax share of an individual is determined by his income and by the community in which he lives. For a citizen of municipality i, with income Y, the tax share will be written $\tau^i(Y)$. Let the demand for a public commodity by a citizen of municipality i be determined by the function $D(Y, \tau^i(Y), n^i, X^i)$ where n^i is the population and X^i a vector of values of variables describing municipality i.

To examine the effects on demand of differences in income within a given community, compute the total derivative of demand with respect to income. This is $dD/dY = \partial D/\partial Y + (\partial D/\partial\tau)(\partial\tau/\partial Y)$. Transforming this result into elasticity form, we have $(Y/D)(dD/dY) = \epsilon + \delta\xi$ where ϵ is the ordinary income elasticity of demand, δ is the price elasticity of demand, and $\xi \equiv (Y/\tau^i)(\partial\tau^i/\partial Y)$ is the elasticity of the tax share with respect to income. If for all values of Y, $\epsilon + \delta\xi > 0$, then the higher a citizen's income the more he will demand of the municipal commodity, if $\epsilon + \delta\xi < 0$ the quantity demanded decreases with income.[10] In either case, quantity demanded is a monotone function of income. It would then follow that the quantity demanded by the citizen with the median income in any municipality is also the median quantity demanded in that municipality. In such cases if we are willing to assume that the municipality is in Bowen equilibrium, the estimating procedure suggested in the previous section will give us reasonable estimates of price and income elasticities of demand.

If, however, $\epsilon + \delta\xi$ is positive for some levels of income and negative for others, the quantity demanded will not be a monotone function of income. In this case the median quantity demanded will not in general be the quantity demanded by the consumer with the median income.

For example, in Figure 1, if the incomes of half the population lie in the intervals OA and BC while the other half of the population has incomes in the interval AB, then the Bowen equilibrium quantity will be OE rather than OD, which is the quantity desired by the consumer with median

[9] It is generally quite difficult to assign the burden of nonproperty taxes. In the case of Pennsylvania there is a municipal income tax for many cities. We have included this tax in our computations of tax shares.

[10] These notions are also introduced by Buchanan (1964).

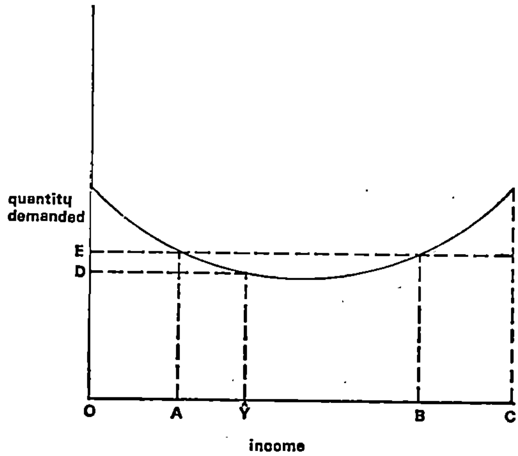


FIGURE 1

income $\hat{Y}$. Obviously, with alternative tax structures, a wide variety of cases can be constructed in which the quantity selected exceeds or falls short of the quantity demanded by the citizen with median income. If there were frequent and substantial variations of this kind, our procedure would not be expected to give reasonable estimates of the income elasticity of demand.

Similar difficulties arise if there are differences in tastes or in tax share functions for different segments of the population of a municipality. Suppose, for example, that renters believe (whether rightly or wrongly) that they do not pay the entire property tax on their housing. Their perceived tax shares will then be lower than those paid by home owners. One would expect them to vote for more public services than their home-owning counterparts. The half of the population desiring the larger amount will consist of proportionally more renters than home-owners. There is again no reason to expect the median quantity desired to be the quantity desired by the consumer with the median income. In general, the Bowen equilibrium quantity will depend on the shape of the entire income distribution functions of each group.

Here, however, things are not quite as chaotic as they may seem. One can make perhaps not grossly unrealistic assumptions which would allow him to use the estimating procedure previously outlined. Matters are greatly simplified if we make certain regularity assumptions about income distributions. We present a theorem which allows us to use the estimating procedure of the previous section even though there may be differences in tastes or in tax share schedules within communities.

Consider a set M of cities. Let P^i be the set of consumers in city i where $i \in M$. Let $\{P_1^i, \ldots, P_n^i\}$ be a partition of P^i, where λ_j^i is the ratio of the number of members of P_j^i to the number of members of P^i. (For each $i \in M$, $\sum_{j=1}^n \lambda_j^i = 1$.) Let F_j^i be the cumulative income distribution function for P_j^i. Thus, $F_j^i(Y)$ is the ratio of the number of members of P_j^i with incomes not exceeding Y to the total number of members of P_j^i.

DEFINITION: *Income distributions are said to be proportional in M (with respect to the partition $\{P_1^i, \ldots, P_n^i\}$) if there exist functions F_j, such that for each $i \in M$ there is a real number $k_i > 0$ where, for all $j = 1, \ldots, n$, $F_j^i(Y) \equiv F_j(k_i Y)$.*

The following remark may help to give intuitive meaning to this definition. Suppose that in municipality 1, the cumulative income distributions for the subsets $P_1^1, \ldots, P_n^1$, are $F_1^1, \ldots, F_n^1$. (These functions need not bear any similarity to each other.) Suppose that income distributions in any other city $i \in M$ are the same as the income distributions in city 1 would be if the income of every citizen of city 1 were multiplied by some number k_i. Income distributions would then be proportional in M.

We can now state Theorem 1. A proof is supplied in the Appendix.

THEOREM 1: *Suppose that for each city*

$i \in M$, the population is partitioned as in the above discussion. Let income distributions be proportional in M with respect to the partition $\{P_1^i, \ldots, P_n^i\}$. Let the tax shares of members of P_j^i with income Y_i to be determined by a function of the form $\tau_j^i(Y) = \tau^i \tau_j Y^\xi$ (where τ^i and τ_j are constants depending on an individual's community and his subset). Let the demands for a public good of a member of P_j^i with income Y be determined by a function of the form $c_j f(X^i)[\tau_j^i(Y)]^\delta Y^\epsilon$. Suppose also that income distribution functions satisfy certain continuity assumptions (which are made explicit in the proof) and that $\epsilon + \xi\delta \neq 0$. Then the median quantity, $\hat{E}^i$, of a public good demanded by citizens of i is determined by a function of the form

$$\hat{E}_i = h(\lambda_1^i, \ldots, \lambda_n^i) c_1 f(X^i) [\hat{\tau}_1^i]^\delta [\hat{Y}^i]^\epsilon$$

where h is a continuous real valued function of the vector $(\lambda_1^i, \ldots, \lambda_n^i)$, $\hat{Y}^i$ is the median income for the entire city i, and $\hat{\tau}_1^i$ is the tax share of a member of P_1^i whose income is $\hat{Y}^i$.

The point of the theorem is that it allows us to use $\hat{Y}^i$ as the only variable describing income in city i and to use the coefficient ϵ as an estimate of income elasticity. Without the assumption of proportionality of income distributions, the quantity $\hat{E}^i$ would generally depend on the detailed structure of income distribution in each subset, P_j^i.

Nothing is said here about the functional form of h, which in general will be rather complicated. However, if we are willing to approximate the effect of the λ's as either linear or *log*-linear, we can estimate the income and price elasticities from cross-sectional observations of the expenditures of municipalities, their median incomes, and the tax shares of consumers of one type with the median income for the municipality. It is this result which justifies, for example, our use of "percent owner occupied" as an independent variable, our use of the tax share of a home-owner with the community median income as the tax share variable, and median income for the community as the income variable.

IV. Results of the Regressions

We were pleased to discover that our estimates of income elasticity were usually significant and positive and the estimates of price elasticity were usually significant and negative. In no case was an estimated elasticity coefficient both significant and of perverse sign. The results for the individual states are reported in Tables 1–3. Table 4 reports the results when all observations are pooled. For these regressions, dummy variables were used to allow the regression line a different intercept for each state. Although the results for the separate states are roughly similar, and there is considerable regularity in the signs of coefficients, the F-test would lead us to reject the hypothesis that mean square errors of the estimates are reduced by pooling the data. (See C. Toro-Viscarrondo and T. Wallace.) Some of the states could be combined successfully according to this criterion. However, rather than apply *ad hoc* methods, we would prefer to defer further efforts at pooling data until better measurement and perhaps better specification of variables is made. Ideally, we would hope that with proper measurement, it will be discovered that the coefficients of our dependent variables are not significantly different between the states. For the present we must settle for the rough similarity which we do find. It may nevertheless be convenient occasionally to use our estimates of the coefficients from the pooled regressions as crude summaries of our results.

Estimates of γ, the crowding parameter, were made by computing the ratio $\alpha/(1+\delta)$ where α and δ are the estimated elasticities of expenditures with respect to population and tax share, respectively. For most

TABLE 1—DETERMINANTS OF CURRENT GENERAL EXPENDITURES OF MUNICIPALITIES IN 1962[a]
(excluding education and welfare)

Coefficients of other variables	California	Illinois[b]	Michigan	Minnesota	Missouri	New Jersey	New York	Ohio	Pennsylvania	Wisconsin
Income elasticity ϵ	0.28	1.73*	0.88*	1.29*	1.65*	0.79*	1.03*	0.80*	0.37	0.16
	0.17	*0.44*	*0.22*	*0.38*	*0.46*	*0.14*	*0.17*	*0.19*	*0.26*	*0.37*
Tax share elasticity δ	−0.39*	−0.29*	−0.41*	−0.25*	−0.25	−0.13*	−0.50*	−0.21*	−0.15	−0.01
	0.08	*0.08*	*0.13*	*0.11*	*0.31*	*0.05*	*0.14*	*0.10*	*0.10*	*0.07*
Population elasticity α	0.67*	0.46*	0.58*	0.69*	0.94*	0.97*	0.75*	0.85*	0.99*	1.00*
	0.09	*0.12*	*0.14*	*0.19*	*0.33*	*0.06*	*0.15*	*0.11*	*0.12*	*0.11*
Crowding parameter $\gamma=(\alpha/(1+\delta))$	1.10	0.65**	0.98	0.92	1.25	1.11**	1.50**	1.08	1.16**	1.01
Percent population change (1950–60)	−0.06*	−0.10	−0.00	−0.03	−0.03	0.02	−0.03	−0.08*	0.03	−0.07*
	0.01	*0.07*	*0.02*	*0.02*	*0.03*	*0.03*	*0.17*	*0.02*	*0.02*	*0.03*
Employment residential ratio	0.11*	−0.08	0.22	−0.13	−0.02	0.24*	0.10	0.05	0.22*	0.12
	0.03	*0.12*	*0.12*	*0.11*	*0.36*	*0.04*	*0.05*	*0.05*	*0.07*	*0.15*
Percent owner occupied	−0.07	−1.56	−1.71*	−3.21*	0.05	−0.78*	−1.68*	−0.58	−0.35	0.35
	0.28	*0.82*	*0.52*	*1.01*	*1.15*	*0.24*	*0.45*	*0.39*	*0.52*	*0.67*
Percent nonwhite	0.45	1.95*	−0.20	12.79	1.80	1.35*	−0.19	0.75	0.48	3.63
	0.46	*0.92*	*0.54*	*10.92*	*2.23*	*0.33*	*0.72*	*0.89*	*0.57*	*2.58*
Density	−0.07	0.05	−0.02	0.03	−0.30	−0.03	−0.22*	−0.15*	0.08	0.02
	0.05	*0.12*	*0.07*	*0.08*	*0.17*	*0.04*	*0.05*	*0.07*	*0.06*	*0.06*
Percent population 65+	1.69*	4.21	−1.35	1.49	8.85*	1.32	3.70*	0.37	1.10	−0.57
	0.81	*2.68*	*1.76*	*2.01*	*2.67*	*0.96*	*1.47*	*1.15*	*1.72*	*2.37*
Percent living in same house (1955–60)	−1.21*	−.57	0.90	−1.48*	−2.88*	0.08	−1.27	−0.79	−0.88	−0.60
	0.35	*0.88*	*0.58*	*0.66*	*1.09*	*0.43*	*0.64*	*0.51*	*0.56*	*0.75*
Intercept	3.02	−9.81	−1.48	−6.08	−11.45	−7.01	−1.68	−4.09	−4.56	−2.68
Number of observations	160	62	70	36	33	120	74	106	124	41
R^2	.89	.80	.91	.89	.89	.94	.94	.87	.81	.96

[a] Values in italics are the standard errors of the coefficients.
[b] Excluding Cook County.
* Indicates a coefficient that is significant at the 95% confidence level.
** Indicates a value of γ that is significantly different from 1 at the 95% confidence level.

TABLE 2—DETERMINANTS OF MUNICIPAL EXPENDITURES ON POLICE IN 1962[a]

Coefficients of other variables	California	Illinois[b]	Michigan	Minnesota	Missouri	New Jersey	New York	Ohio	Pennsylvania	Wisconsin
Income elasticity ϵ	0.26	1.89*	0.54	0.80	1.04*	0.94*	1.78*	1.08*	0.99*	0.95
	0.16	*0.34*	*0.62*	*0.41*	*0.28*	*0.14*	*0.77*	*0.17*	*0.41*	*0.86*
Tax share elasticity δ	−0.25*	−0.19*	−0.76*	−0.13	−0.29	−0.15*	−0.31	−0.18*	−0.36*	−0.15
	0.07	*0.06*	*0.36*	*0.13*	*0.19*	*0.05*	*0.64*	*0.09*	*0.15*	*0.16*
Population elasticity α	0.75*	0.56*	0.26	0.83*	0.74*	0.81*	1.02	0.77*	0.67*	0.75*
	0.08	*0.09*	*0.40*	*0.20*	*0.21*	*0.07*	*0.70*	*0.10*	*0.18*	*0.25*
Crowding parameter $\gamma=(\alpha/(1+\delta))$	1.00	0.69**	1.08	0.95	1.04	0.95	1.48	0.94	1.05	0.88
Percent population change (1950–60)	−0.04*	−0.06	−0.07	−0.04	0.01	0.04	−2.73*	−0.01	0.06	−0.14*
	0.01	*0.05*	*0.05*	*0.02*	*0.02*	*0.03*	*0.77*	*0.02*	*0.03*	*0.06*
Employment residential ratio	0.08*	−0.04	−0.27	−0.10	0.06	0.21*	−0.14	0.06	0.20	0.19
	0.03	*0.09*	*0.33*	*0.12*	*0.22*	*0.04*	*0.23*	*0.04*	*0.12*	*0.36*
Percent owner occupied	−0.19	−2.25*	−0.51	−3.61*	−0.81	−0.65*	−1.16	−0.59	0.38	−0.08
	0.26	*0.62*	*1.47*	*1.10*	*0.71*	*0.24*	*2.07*	*0.35*	*0.81*	*1.57*
Percent nonwhite	0.69	1.53*	0.13	0.97	0.87	1.21*	2.32	1.43	0.04	−1.55
	0.42	*0.70*	*1.52*	*11.94*	*1.38*	*0.34*	*3.30*	*0.80*	*0.89*	*6.07*
Density	−0.02	0.01	0.03	0.05	−0.09	0.07	−0.46*	0.05	0.55*	0.20
	0.04	*0.09*	*0.21*	*0.08*	*0.10*	*0.04*	*0.23*	*0.07*	*0.09*	*0.15*
Percent population 65+	0.99	−0.97	1.21	−0.85	1.16	0.55	10.38	−1.39	0.77	2.47
	0.74	*2.03*	*4.93*	*2.20*	*1.66*	*0.98*	*6.77*	*1.03*	*2.68*	*5.58*
Percent living in same house (1955–60)	−0.70*	0.53	0.06	−0.74	−0.33	0.47	−9.11*	0.25	−1.50	−2.39
	0.32	*0.67*	*1.64*	*0.72*	*0.62*	*0.44*	*2.96*	*0.46*	*0.87*	*1.76*
Intercept	−0.80	−13.70	4.36	−5.59	−7.12	−9.24	−8.58	−9.90	−11.31	−9.02
Number of observations	160	62	70	36	33	120	74	106	124	41
R^2	.89	.87	.57	.88	.93	.92	.69	.89	.69	.88

[a] Values in italics are the standard errors of the coefficients.
[b] Excluding Cook County.
* Indicates a coefficient that is significant at the 95% confidence level.
** Indicates a value of γ that is significantly different from 1 at the 95% confidence level.

TABLE 3—DETERMINANTS OF MUNICIPAL EXPENDITURES ON PARKS AND RECREATION IN 1962[a]

Coefficients of other variables	California	Illinois[b]	Michigan	Minnesota	Missouri	New Jersey	New York	Ohio	Pennsylvania	Wisconsin
Income elasticity ϵ	0.10	3.54	1.25	2.19*	2.81*	2.28*	1.74*	1.50*	1.76*	−1.87
	0.44	*1.98*	*0.66*	*0.83*	*0.94*	*0.54*	*0.64*	*0.55*	*0.80*	*1.18*
Tax share elasticity δ	−0.68*	0.01	0.25	−0.39	−0.24	−0.00	−0.81	−0.49	0.02	−0.13
	0.20	*0.37*	*0.38*	*0.25*	*0.63*	*0.17*	*0.53*	*0.28*	*0.29*	*0.22*
Population elasticity α	0.67*	1.01	1.49*	0.74	1.11	1.28*	0.69	0.60	1.72*	1.19*
	0.23	*0.52*	*0.43*	*0.41*	*0.69*	*0.25*	*0.58*	*0.31*	*0.36*	*0.35*
Crowding parameter $\gamma=(\alpha/(1+\delta))$	2.09**	1.00	1.19	1.21	1.47	1.28	3.63**	1.18	1.69**	1.37
Percent population change (1950–60)	−0.11*	0.33	0.01	−0.02	−0.22*	−0.15	1.42*	−0.08	0.05	−0.44*
	0.02	*0.30*	*0.05*	*0.04*	*0.07*	*0.12*	*0.64*	*0.06*	*0.07*	*0.08*
Employment residential ratio	0.08	0.05	1.25*	−0.19	1.22	0.41*	0.35	−0.10	0.83*	−0.83
	0.09	*0.51*	*0.35*	*0.24*	*0.75*	*0.16*	*0.19*	*0.14*	*0.23*	*0.49*
Percent owner occupied	0.00	−4.03	0.25	−4.90*	4.08	−2.06*	−3.48*	−0.27	0.23	2.46
	0.73	*3.64*	*1.56*	*2.20*	*2.30*	*0.91*	*1.72*	*1.13*	*1.58*	*2.15*
Percent nonwhite	−2.06	−6.06	−0.33	−1.69	−0.57	2.31	1.66	3.03	−0.23	−5.41
	1.19	*4.09*	*1.61*	*23.95*	*4.61*	*1.28*	*2.75*	*2.62*	*1.74*	*8.35*
Density	0.05	0.49	−0.39	0.15	0.26	0.38*	−0.21	−0.31	0.35*	−0.19
	0.13	*0.52*	*0.22*	*0.17*	*0.35*	*0.16*	*0.19*	*0.21*	*0.17*	*0.21*
Percent population 65+	1.22	25.87*	−8.23	8.99*	11.25*	3.92	14.27*	−2.33	7.11	−4.85
	2.10	*11.91*	*5.25*	*4.15*	*5.53*	*3.68*	*5.64*	*3.38*	*5.24*	*7.67*
Percent living in same house (1955–60)	−1.00	−0.92	0.81	−2.38	−9.27	−0.57	−3.14	−0.90	−4.77*	−5.29*
	0.92	*3.93*	*1.74*	*1.45*	*2.06*	*1.65*	*2.46*	*1.49*	*1.70*	*2.42*
Intercept	3.84	−40.84	−19.33	−15.99	−31.62	−23.20	−8.03	−12.99	−29.73	15.87
Number of observations	160	62	70	36	33	120	74	106	124	41
R^2	.67	.35	.63	.79	.83	.63	.67	.56	.60	.86

[a] Values in italics are the standard errors of the coefficients.
[b] Excluding Cook County.
* Indicates a coefficient that is significant at the 95% confidence level.
** Indicates a value of γ that is significantly different from 1 at the 95% confidence level.

states, the estimates of γ for police expenditures and general expenditures are strikingly close to unity. We applied an F-test to determine whether to reject the hypothesis that the linear restriction $\alpha=1+\delta$ applies. In all states but Illinois and in the pooled regression, the test strongly suggests that the null hypothesis should be accepted for police expenditures. For general expenditures, we would have to reject the null hypothesis at the 5 percent level for about half of the states and for the pooled regression. However, in all cases except Illinois we would reject the hypothesis that $\gamma \leq 1$. The test also suggests that for parks and recreation expenditures, γ is one or greater.

Some economic implications of the values of the price and income elasticities and the coefficient, γ, will be discussed in the next section.

A few speculations about the effects of some of the auxiliary variables may be justified by the degree of regularity with which the same signs appear. The differences in their sizes and the possibility of correlation with other more "truly explanatory" variables does suggest caution in making causal interpretations. The reader is, of course, invited to make his own speculations.

Percent owner occupied: As argued earlier in this paper, it may be that renters do not believe that they pay the entire property tax on their housing, and tend to vote for more public expenditures than homeowners with the same income. If this is the case, one would expect the negative coefficient which we find for this variable.

Employment-Residential ratio: Communities with high values for this variable would tend to have a large amount of com-

TABLE 4—DETERMINANTS OF MUNICIPAL EXPENDITURES, 1962
ALL OBSERVATIONS POOLED[a]

	General Expenditures	Police Expenditures	Parks & Recreation
Income elasticity ϵ	0.64*	0.71*	1.32*
	0.07	*0.13*	*0.22*
Tax share elasticity δ	−0.23*	−0.25*	−0.19*
	0.03	*0.05*	*0.08*
Population elasticity α	0.84*	0.80*	1.17*
	0.03	*0.06*	*0.11*
Crowding parameter $\gamma=(\alpha/(1+\delta))$	1.09**	1.07	1.44**
Percent population change (1950–60)	−0.04*	−0.04*	−0.08*
	0.01	*0.01*	*0.02*
Employment residential ratio	0.12*	0.01	0.24*
	0.02	*0.04*	*0.06*
Percent owner occupied	−0.77*	−1.12*	−0.78
	0.13	*0.25*	*0.42*
Percent nonwhite	0.84*	0.90*	−0.20
	0.19	*0.36*	*0.60*
Density	−0.07*	0.01	−0.02
	0.02	*0.04*	*0.06*
Percent population 65+	1.75*	1.27	4.94*
	0.45	*0.85*	*1.43*
Percent living in same house (1955–60)	−0.65*	−0.77*	−1.99*
	0.17	*0.32*	*0.53*

[a] Values in italics are the standard errors of the coefficients.
* Indicates a coefficient that is significant at the 95 percent confidence level.
** Indicates a value of γ that is significantly different from 1 at the 95 percent level.

mercial or industrial activity. It may be that larger amounts of public services must be provided in order to attract and retain such activity. This might explain the positive effect which we find.

Population change: Cities which have grown rapidly may not yet have achieved a political equilibrium. Time is required for a rapidly growing city to gain a political consensus for expanding public services. Inertial effects may also result in larger expenditures in places with declining population. This might explain our negative coefficient.

Percent of population over 65: The life cycle hypothesis would predict that persons over 65 years of age tend to spend a larger portion of their current income on current consumption than younger people. If demand for public goods as a fraction of total demand does not diminish with age then one would expect an aged person to demand a larger quantity of public goods than a younger person with the same income and tax share. The effect of this variable would then be positive, which is what we usually find.

The variables, percent nonwhite, percent who lived in the same house from 1955 to 1960, and population density, frequently had coefficients significantly different from zero. We have no particularly compelling explanation for the pattern of signs which emerges. Although these variables are included in the regressions reported, the size and significance of the other coefficients are not substantially altered when we exclude them.

TABLE 5

If 95 percent of the measured values $\hat{T}$ lie in the interval [,] where T^* is the "true" value of median tax share	$[.5T^*, 2T^*]$	$[.57T^*, 1.75T^*]$	$[.66T^*, 1.5T^*]$	$[.75T^*, 1.33T^*]$	$[.9T^*, 1.11T^*]$	No error
then the maximum likelihood estimate of δ is	$-.83$	$-.44$	$-.30$	$-.26$	$-.24$	$-.23$

We experimented with the use of some other variables, including population per household, median school years completed, percent of population less than 18 years old, whether a municipality was a suburb or not, percent of the population with annual incomes below \$3,000 and percent of the population with annual incomes exceeding \$10,000. Each of these variables proved to have statistically insignificant effects in all or nearly all of our regressions.

The fact that the variables representing percentages of populations at the extreme ends of the income distribution were almost always insignificant provides mild evidence in favor of our hypothesis that the consumer with the median income also demands the median quantity of public goods. It is of some interest to observe however, that for California, the percentage of the application with income under \$3,000 has a significant and positive effect on demand. When this variable is included, the coefficient of median income for California rises from values considerably lower than those displayed for other states to values quite similar to the others.

It was thought that perhaps our estimates would be overwhelmed by the effects of population size. In particular there is a strong negative correlation between population and tax share (because in larger cities taxes are divided among more persons). It also seemed possible that the effects of some of our variables might be substantially different in cities of different size, either due to differences in political structure or uncaptured social differences. We therefore stratified our sample into nine population subgroups. Again, our coefficients were almost all reasonable and fairly similar between groups. For no population subgroup were we able to reject the hypothesis that $\gamma=1$.

As mentioned above, we computed maximum likelihood estimates for the regression coefficients under some alternative assumptions about errors in the measurement of tax share. The only coefficient which was substantially affected was the coefficient of tax share itself. In Table 5 we report the maximum likelihood estimate of the tax share elasticity δ where there is a lognormally distributed measurement error with alternative variances.[11] These values apply to the pooled regression on general expenditures. The effects on the other regressions were similar.

V. Some Implications of the Results

The reasonable values which we find for our coefficients lend at least some support to the hypotheses which we make about political behavior. These hypotheses in turn provide a useful tool for the estimation of the effects of income, price, city size, and other social variables on demand

[11] We follow the maximum likelihood procedure suggested by John Johnston, pp. 168–70.

for public commodities. The reader would, of course, do well to maintain a degree of skepticism about the values of the coefficients which we find. Strong assumptions were made to derive the estimates. It is possible that more precise measurement of the variables and additional tests of the assumptions would result in quite different estimates of demand functions.

Here we examine some implications of the parameter values which we find. The reader is free to apply the analysis below to the implications of any alternative estimates which he prefers.

Suppose that the only real property holdings of voters are their dwellings and that the tax share of a voter is proportional to his holdings of real property. The most commonly accepted estimates of the income elasticity of demand for housing appear to be between 1 and 1.3.[12] This would imply an elasticity, ξ, of tax share with respect to income in the same range. For most of our regressions (exceptions are California and Pennsylvania) we estimate the income elasticity, ϵ, to be substantially greater than $-(1.3)\delta$. Thus, we would expect that $\epsilon+\xi\delta>0$. According to our earlier discussion one would expect that in most cities, citizens with incomes exceeding the median would vote for increased expenditures and citizens with lower incomes would vote against increased expenditures.[13]

The Lindahl solution for allocation of public goods has predictive and normative properties which are attractive to many economists. Under quite reasonable assumptions, Lindahl equilibrium exists, is Pareto optimal, and is in the core.[14] In Lindahl equilibrium, tax shares are arranged in such a way that all consumers demand the same quantity of each public commodity. It has frequently been pointed out that if an individual's stated preferences for public commodities influence his tax share, it is usually in his interest to misrepresent his preferences. This difficulty would seem to greatly reduce the practical usefulness of the Lindahl system as an allocative device. As our paper suggests, it may nevertheless be possible to gain approximate information about the preferences of broadly defined social groups if one observes the voting behavior of individuals under circumstances where it is reasonable to believe that an individual's expressed desires concerning levels of expenditure do not influence his tax share.

A Bowen equilibrium is not, in general, Pareto optimal. For "most" Bowen equilibria one could find an alternative distribution of the tax burden and alternative levels of public expenditure which would result in a Pareto-superior allocation. One might wish to ask whether a certain Bowen equilibrium results in "too little" or "too much" public goods. This question is not generally answerable on the basis simply of the Pareto criterion. As Paul Samuelson stresses, there are an infinity of Pareto optimal allocations corresponding to different points on the "utility possibility" frontier. Generally there will be different quantities of public goods for different Pareto optimal allocations. Some of these quantities may be larger and some smaller than the Bowen equilibrium quantity considered.

A more determinate normative criterion would be to compare a given Bowen equilibrium to the Lindahl equilibrium corresponding to the initial distribution of income. If voters of all incomes are to agree about the quantity of public goods to be supplied, the tax schedule must be

[12] See for example the papers by Richard Muth and Frank de Leeuw.

[13] An interesting piece of independent evidence indicating the same conclusion is provided by Harlan Hahn and W. B. Shephard's study of precinct voting in municipal referenda.

[14] See the papers of Erik Lindahl, Duncan Foley, and Bergstrom (1970).

such that $\epsilon+\xi\delta=0$.[15] For a community with known values of $-\epsilon/\delta$ and ξ, we could compute both the Lindahl equilibrium and the Bowen equilibrium if we had full knowledge of the income distribution. It is in general not possible with knowledge only of $-\epsilon/\delta$ and ξ to determine in which case the quantity of public goods is larger.

If, however, we assume that the elasticities, ξ and $-\epsilon/\delta$ are constant over all incomes, we can provide a partial answer to this question given only limited information about the income distribution. In particular, in either Bowen equilibrium or in Lindahl equilibrium the quantity supplied will be equal to the quantity demanded by the citizen with median income. To determine in which case he demands more, we need only to determine whether his tax share is lower in Lindahl equilibrium or in Bowen equilibrium. Where price elasticities are negative, the Lindahl equilibrium quantity will exceed the Bowen equilibrium quantity if, and only if, the tax share of the consumer with median income is lower in Lindahl equilibrium. In another paper (1973), Bergstrom has shown that if mean income in a community exceeds median income, if $-\epsilon/\delta>1$ and $\xi<-\xi/\delta$, the tax share of the citizen with median income will be lower in Lindahl equilibrium than in Bowen equilibrium. Our estimates suggest that each of these assumptions is likely to hold. It then follows that in the communities we observe, less than the Lindahl equilibrium quantity is provided. One could argue that this result provides some evidence to support John Kenneth Galbraith's contention that in reaching a "truce on income distribution" communities arrive at a quantity of public goods which is less than optimal. On the other hand, our estimates of income elasticities of less than unity for general municipal expenditures would seem "anti-Galbraithian."[16]

Our results suggest that the crowding parameter γ generally has a value of one or greater for cities with populations between 10,000 and 150,000. One could interpret this to mean that as the size of municipalities increases, the advantages of sharing the cost of public services among more persons are countervailed by the costs of sharing the services among more persons. An alternative statement of this proposition would be that over the range of city sizes which we studied, there appear to be no economies of scale to larger municipalities in the provision of public goods. It should, of course, be understood that this statement says nothing about the question of whether there are economies of scale to production of publicly supplied goods measured in some choice of physical units, but relates only to net effects of city size after crowding phenomena are accounted for.

One might reasonably ask why, if there are not increasing returns in the municipal provision of the goods and services which we study, is their provision in the public domain. The answer may be that there are substantial scale economies in the collective provision of these commodities

[15] Where some of the burdens of taxes and benefits of public goods are shared by noncitizens, we might choose to define Lindahl equilibrium in a municipality either as a situation in which there is agreement among all citizens or among all interested parties. If we take the former approach, Lindahl equilibrium is not necessarily Pareto optimal from the viewpoint of the national economy. Even here, one would expect that competition among municipalities for local industry would tend to result in efficient provision of public services. In any case, Lindahl equilibrium does require that there be agreement among citizens of all incomes. Thus, the requirements for income elasticities of tax shares remain unaltered.

[16] Galbraith's position, however, rests at least in part on the claim that expressed demands are a distortion of underlying "true" preferences since greater advertising of private goods reduces relative expressed demands. Hence he might choose to deny any normative significance for the demand functions which we estimate.

for units of smaller size than the cities which we consider.[17] This could provide sufficient reason that they be collectively provided. It may be that once some critical size is achieved, city size becomes largely a matter of indifference from an efficiency viewpoint over a fairly large range of sizes. Some corroborating evidence is provided by the fact that in metropolitan areas, contiguous municipalities frequently differ substantially and seemingly arbitrarily in size.

Appendix

Part I: Individual Errors of Perception

We introduce three lemmas, the first two of which were proved by Mosteller.

LEMMA 1: *If a random sample of size n is drawn from a probability distribution, f, which is continuous and nonvanishing in a neighborhood of its median, $\tilde{X}$, then the sample median $\hat{X}(n)$ is asymptotically normally distributed (as n becomes large) with expected value $\tilde{X}$ and variance $\sigma^2(n)=1/(4n[f(\tilde{X})]^2)$ where $\text{limit}_{n\to\infty}\ \sigma^2(n)=0$.*

LEMMA 2: *Let the random variable $\theta(n)$ be a function of the values $X_1, \ldots, X_n$ in a random sample of size n drawn from a probability distribution f. If $\theta(n)$ is asymptotically normally distributed and $\lim_{n\to\infty}\ [E\theta(n)]=\bar{\theta}$ where the asymptotic variance of $\theta(n)$ is zero, then if*

$$g[\theta(n)]$$

is a continuous real-valued function with a nonvanishing derivative in a neighborhood of $\bar{\theta}$, $g[\theta(n)]$ is an asymptotically normally distributed random variable such that

$$\lim_{n\to\infty} E[g(\theta(n))] = g(\bar{\theta})$$

and the asymptotic variance of g is zero.

[17] In the case of municipal courts and some policing functions it might be argued that these are social reasons for public provision quite distinct from economies of scale.

LEMMA 3: *If $X_1, \ldots, X_n, Y_1, \ldots, Y_n$ are symmetrically distributed and mutually independent random variables and if $Z_i=X_i+Y_i$ for $i=1, \ldots, n$, then $E(\hat{Z})=E(\hat{X})+E(\hat{Y})$ where $\hat{}$ denotes the sample median of a random variable.*

PROOF:

Since X, Y, and Z are symmetrically distributed, if we let a bar denote the sample mean, then

$$E(\hat{Z})=E(\overline{Z})=E(\overline{X})+E(\overline{Y})=E(\hat{X})+E(\hat{Y})$$

THEOREM: *In a municipality with n citizens, let the demand of consumer i with income Y_i for a public good be a random variable*

$$D_i = c[\tau^*(Y_i)]^{\delta}Y_i^{\epsilon}$$

Let

$$\tau(Y_i) = tY_i^{\xi}$$

be the "true" tax share and $\tau^*(Y_i)=Z_i\tau(Y_i)$ be his perceived tax share where Z_i is a random variable. Suppose that the values of Y and Z for the population of the city are determined by a random sample of size n drawn from a population in which $\log Z$ and $\log Y$ are independently and symmetrically distributed and where $E(\log Z)=0$ and $E(\log Y)=\log \hat{Y}$ where $\hat{Y}$ is the observed median income in the municipality. Then if $\hat{D}$ is the median quantity demanded by citizens of the municipality, $\log \hat{D}=\log c+\delta \log \tau(\hat{Y})+\epsilon \log \hat{Y}+\gamma$ where γ is asymptotically normally distributed with mean of zero and asymptotic standard deviation equal to zero.

PROOF:

From the demand equation we have,

$$\log D_i = (\delta\xi + \epsilon) \log Y_i + \delta \log Z_i$$

Applying Lemma 3, we have

$$E(\widehat{\log D}) = E[(\delta\xi + \epsilon) \log Y] + E(\widehat{\delta \log Z}) + \log c$$

where the hat denotes the sample median. But for any monotonic function f, $\widehat{f(x)}=f(\hat{x})$.

Hence,

$$E(\log \hat{D}) = E[(\delta\xi + \epsilon) \log \hat{Y}) + E(\delta \log \hat{Z}) + \log c$$

Lemma 2 implies that

$$\lim_{n\to\infty} E(\log \hat{D}) = (\delta\xi + \epsilon) \log [\lim_{n\to\infty} E(\hat{Y})] + \log [\lim_{n\to\infty} E(\hat{Z})] + \log c$$

Applying Lemma 1 and our assumptions about the distributions of Y and Z, we have

$$\lim_{n\to\infty} E(\log \hat{D}) = (\delta\xi + \epsilon) \log \hat{Y} + \log c = \log c + \delta \log \tau(\hat{Y}) + \epsilon \log \hat{Y}$$

By Lemma 1, the asymptotic variance of $\log \hat{D}$ is zero. The theorem follows immediately.

This result provides us with a partial justification for presuming that individual errors of perception will have no statistically unpleasant effects if the population is very large. Two problems remain. The Theorem assumes that $\log Y$ is symmetrically distributed. If this is not the case, then Lemma 3 does not apply and $\epsilon \log \hat{Y} + \delta \log \tau(\hat{Y}) + \log c$ may be a biased estimate of $\log \hat{D}$. This would not be a severe problem unless the asymmetry is substantial and highly variable from city to city. If, for example, income distributions are proportional between cities (as defined in the text) then the income distribution functions in different cities are simply translations of each other. In this case, the bias is the same for each city. The bias would affect the regression intercept but not the estimates of elasticities.

It also should be demonstrated that for cities as large as those which we consider, the asymptotic results are closely approximated. To find the magnitudes of these effects we can use the result that the variance of the median of a random sample of size n drawn from a normal distribution with variance σ^2 will be no larger than $(1/n)(2/\pi)\sigma^2$. (See J. Hodges and E. Lehmann.) If we postulate reasonable values for our parameters, for a community of 2,000 or more families, even very substantial variances of individual perception error result in a variance of the error term σ discussed above which would have negligible statistical effects.

It might also be noted that it is well that these error terms tend to be small, because they are not "well-behaved" in other ways. Although when the error is that of the actor rather than the observer it may be reasonable to assume that the error term is uncorrelated with the observed value of the variable (unlike the usual assumption when the error is that of the observer), it remains the case that the variance of the error term is likely to be smaller as city size increases. This violates the classical assumptions.

Part II: Proof of Theorem 1

For each $i \in M$ and $j = 1, \ldots, n$, let I_j^i be the closed interval between the maximum and the minimum income in p_j^i. We make the following continuity assumption. (C) For every $i \in M$, the set $\bigcup_{j=1}^n I_j^i$ is a closed interval, and the functions F_j^i, $j = 1, \ldots, n$, are all continuous and strictly increasing on I_j^i.

The cumulative income distribution, F^i, for the entire population P^i is easily seen to be

$$F^i = \sum_{j=1}^n \lambda_j^i F_j^i$$

Assumption C insures that F^i is continuous and strictly increasing on $\bigcup_{j=1}^n I_j$. If this assumption were dispensed with, one could prove a result similar to Theorem 1 which applies to discrete distributions. However, for our purposes, the added realism does not appear to justify the added tedium in the statement and proof of the theorem.

Let $\hat{Y}^i$ be the median income for the population P^i. Then

$$1/2 = F^i(\hat{Y}^i) = \sum_{j=1}^n \lambda_j^i F_j^i(\hat{Y}^i) = \sum_{j=1}^n \lambda_j^i F_j(k_i \hat{Y}^i)$$

It follows that the value of $(\lambda_1^i, \ldots, \lambda_n^i)$ determines $k_i \hat{Y}^i$. In fact since F^i is continuous and strictly increasing on $\bigcup_{j=1}^n I_j^i$, there is a continuous function g such that

$$k_i\hat{Y}^i = g(\lambda_1^i, \ldots, \lambda_n^i)$$

Since $\epsilon+\xi\delta\neq 0$, one can choose for each $i\in M$ and for $j=1, \ldots, n$, a $\tilde{Y}_j^i$ such that if $\hat{E}^i$ is the median quantity demanded in city i,

then

$$\hat{E}^i = c_j f(X^i)(t^i t_j)^\delta [\tilde{Y}_j^i]^{\epsilon+\xi\delta}$$

Thus for $j=1, \ldots n$, $\tilde{Y}_j^i = \rho_j \tilde{Y}_1^i$

where

$$\rho_j = \left[\frac{c_j t_j^\delta}{c_1 t_1^\delta}\right]^{1/(\epsilon+\xi\delta)}$$

Since E is the median quantity demanded, it must be that

$$1/2 = \sum_{j=1}^{n} \lambda_j^i F_j^i(\tilde{Y}_j^i) = \sum_{j=1}^{n} \lambda_j^i F_j^i(\rho_j \tilde{Y}_1^i)$$
$$= \sum_{j=1}^{n} \lambda_j^i F_j(k_i \rho_j \tilde{Y}_1^i)$$
$$= \sum_{j=1}^{n} \lambda_j^i F_j(g(\lambda_1^i, \ldots \lambda_n^i)\rho_j \frac{\tilde{Y}_1^i}{\hat{Y}^i}$$

The latter expression is a continuous strictly increasing function of $\tilde{Y}_1^i/\hat{Y}^i$. It follows that $\tilde{Y}_1^i/\hat{Y}^i$ is determined by the vector $(\lambda_1^i, \ldots \lambda_n^i)$. In fact, there is a continuous function h such that

$$h(\lambda_1^i, \ldots \lambda_n^i) = \left[\frac{\tilde{Y}_1^i}{\hat{Y}^i}\right]^{\epsilon+\xi\delta}$$

But

$$\hat{E}^i = c_1 f(X^i)(t^i t_1)^\delta \tilde{Y}_1^{i\epsilon+\xi\delta}$$
$$= h(\lambda_1^i, \ldots, \lambda_n^i) c_1 f(X^i)(\tau_1^i)^\delta (\hat{Y}^i)^\epsilon$$

REFERENCES

R. Barlow, "Efficiency Aspects of Local School Expenditures," *J. Polit. Econ.*, Sept./Oct. 1970, *78*, 1028–40.

J. Barr and O. Davis, "An Elementary Political and Economic Theory of the Expenditures of Local Governments," *Southern Econ. J.*, Oct. 1966, *33*, 149–65.

T. Bergstrom, *Collective Choice and the Lindahl Allocation Method*, mimeo. 1970.

———, "A Note on Efficient Taxation," *J. Polit. Econ.*, forthcoming.

T. Borcherding and R. Deacon, "The Demand for the Services of Non-Federal Governments," *Amer. Econ. Rev.*, Dec. 1972, *62*, 891–901.

H. Bowen, "The Interpretation of Voting in the Allocation of Economic Resources," *Quart. J. Econ.*, Nov. 1943, *58*, 27–48; reprinted in K. Arrow and T. Scitovsky, eds., *Readings in Welfare Economics*, Homewood 1969.

J. Buchanan and G. Tullock, *The Calculus of Consent*, Ann Arbor 1962.

F. de Leeuw, "The Demand for Housing: *A Review of Cross-Section Evidence*," *Rev. Econ. Statist.*, Feb. 1971, *53*, 1–10.

D. Foley, "Lindahl's Solution and the Core of an Economy with Collective Goods," *Econometrica*, Jan. 1970, *38*, 66–72.

J. Galbraith, *The Affluent Society*, Boston 1958.

H. Hahn and W. B. Shepard, *Public Preferences, Economic Resources, and Governmental Expenditures: A Study of Municipal Referenda*, mimeo. Riverside, California.

G. Hardy, H. Littlewood, and H. Polya, *Inequalities*, 2d ed., London 1952.

J. Hodges and E. Lehmann, "On Medians and Quasi-medians," *J. Amer. Statist. Assn.*, Sept. 1967, *62*, 926–31.

J. Johnston, *Econometric Methods*, New York 1960.

E. Lindahl, "Just Taxation—A Positive Solution," in R. Musgrave and A. Peacock, eds., *Classics in the Theory of Public Finance*, London 1958.

F. Mosteller, "On Some Useful "Inefficient" Statistics," *Ann. Math. Stat.*, Dec. 1946, *17*, 377–408.

R. Muth, *Permanent Income, Instrumental Variables and the Income Elasticity of Housing Demand*, mimeo. 1970.

P. Samuelson, "The Pure Theory of Public Expenditure," *Rev. Econ. Statist.*, Nov. 1954, *36*, 387–89.

C. Toro-Vizcarrondo and T. Wallace, "A Test of the Mean Square Error Criterion in Linear Regression," *J. Amer. Statist. Assn.*, June 1968, *63*, 558–72.

Intermediate Products and the Pure Theory of International Trade: A Neo-Heckscher-Ohlin Framework

By RAVEENDRA N. BATRA AND FRANCISCO R. CASAS*

Economists in general and trade theorists in particular have now come to recognize the importance of intermediate inputs, not merely of primary factors, in the process of production. This is well reflected in the recent studies by Jaroslav Vanek, James Melvin (1969), and Roy Ruffin, among others. The main feature of these models, with the singular exception of Ruffin's, is their treatment of two commodities—traditionally taken to be consumption goods—as both final outputs and intermediate inputs for each other.[1] With such a minor difference, it is not surprising to find that the traditional theorems by T. Rybczynski, Wolfgang Stolper and Paul Samuelson, Eli Heckscher and Bertil Ohlin, etc., hold without any additional qualification in Vanek's model where the final goods also serve as intermediate products. Most of these traditional theorems depend crucially upon the factor-intensity ranking of the traded commodities, and as long as this ranking remains unaltered, as it does in Vanek's interindustry flows framework, these theorems will be valid in spite of the introduction of intermediate products.

Citing Vanek's results, Murray Kemp (1969) has defended the neglect of intermediate products in the earlier trade theory. In this paper, however, we intend to show that most of the traditional trade theorems may not hold without additional provisions if such intermediate products are introduced in the model as are produced solely to serve as inputs in the production of final goods. These we shall term *pure* intermediate products in order to distinguish them from the interindustry flows. Furthermore, in a two-good model where intermediate and final products are identical, one cannot explain the basis of trade in intermediate goods. For this reason, there exists no theory at present which would explain why trade occurs in such goods even though, as Paul Yates has shown, the bulk of international trade is in intermediate products—produced goods, like raw materials, spare parts, etc. which are solely used as inputs in the production of other products. Another purpose of this paper is thus to fill this gap and provide a theory explaining the basis of trade in intermediate products. In the traditional two final commodity trade

* Associate professor of economics, Southern Methodist University, and assistant professor, University of Toronto, respectively. An earlier draft of this paper was presented at the winter meetings of the Econometric Society, New Orleans, December 1971. The paper has benefited so much from the valuable suggestions of Jagdish Bhagwati, George Borts, James Melvin, C. Kuo, an unknown referee, and several participants in the trade session at the Econometric Society meetings and the International Economics Workshop at the University of Western Ontario, that the complexion and style of the present version bear little resemblance with those of the original draft. We gratefully acknowledge the generous suggestions offered by these economists; however, none of them is to be held responsible for any remaining errors. Batra's research was financed by a research grant from the Canada Council.

[1] For other studies which treat intermediate goods identically as final goods, see Paul Samuelson, Ronald McKinnon, Francisco Casas (1972a), and Robert Warne. On the other hand, Chulsoon Khang has analyzed a dynamic model of trade between one final good and an intermediate good.

model, we introduce a third good which is produced by primary factors only to serve as an input in the other two commodities. Such a model has been recently developed by Batra and Ranjit Singh to examine the growth behavior of a two-sector economy.

The discussion will evolve as follows. In Section I we give a complete description of the model. The Rybczynski theorem is then discussed in Section II, while Section III is concerned with the Stolper-Samuelson theorem. Finally, Section IV investigates the validity of the Heckscher-Ohlin theorem.

I. The Model with Intermediate Products

It is assumed that the economy consists of three commodities, two final products (X_1 and X_2), and one intermediate good (X_3) which is produced solely to serve as an input in the production of the final products. There are two primary factors of production, capital (K) and labor (L) which enter into the production process of all three commodities, and remain fully employed. Production functions exhibit constant returns to scale and diminishing returns to factor proportions. Perfect competition in product and factor markets, inelastic factor supplies, perfect factor mobility and factor price flexibility, and nonreversibility of factor intensities at all factor-price ratios are also assumed.

The assumption of full employment of primary factors implies that

$$(1) \qquad C_{L1}X_1 + C_{L2}X_2 + C_{L3}X_3 = \overline{L}$$

$$(2) \qquad C_{K1}X_1 + C_{K2}X_2 + C_{K3}X_3 = \overline{K}$$

where X_i denotes the output of the ith industry, $\overline{L}$ and $\overline{K}$ the fixed supplies of primary factors, and C_{Li} and C_{Ki}, respectively, the labor-output and capital-output ratios in the ith productive sector ($i=1, 2, 3$).

Let a_i denote the requirement of the material input, X_3, per unit of the ith final commodity, that is, $a_i = X_{3i}/X_i$. Following Samuelson, Vanek, and others who have worked with input-output models, a_i will be assumed to be constant. Furthermore, from our assumption that X_3 is used only as an input for the two final goods,

$$(3) \qquad X_3 = D_3 = a_1X_1 + a_2X_2$$

where D_3 is the demand for the intermediate product.

Turning now to the pricing of goods and factors, it is readily seen that with perfect competition in all markets and with constant returns to scale, the price of each commodity equals the sum of factor payments per unit of output, that is,

$$(4) \qquad C_{L1}w + C_{K1}r + a_1p_3 = p_1$$

$$(5) \qquad C_{L2}w + C_{K2}r + a_2p_3 = p_2$$

$$(6) \qquad C_{L3}w + C_{K3}r = p_3$$

where w and r denote, respectively, the wage rate and the rental on capital, and p_i is the price of the ith commodity. For the sake of simplicity, we shall assume hereafter that X_1 is the numeraire, so that $p_1 = 1$, and p_2 and p_3 will denote, respectively, the relative price of X_2 and X_3 in terms of X_1.

To complete the supply side of the model, we need only explain the determination of the various production coefficients. As stated earlier, the intermediate input-output coefficients, a_i, are assumed constant. As far as the labor-output and capital-output coefficients are concerned, they depend on factor prices. In particular, if σ_i is the elasticity of substitution between labor and capital in the ith sector, i.e., if $\sigma_i = [d(K_i/L_i)/d(w/r)] \cdot [(w/r)/(K_i/L_i)]$, then with linearly homogeneous production functions we may write

$$(7) \qquad \sigma_i = \frac{\dfrac{dC_{Ki}}{C_{Ki}} - \dfrac{dC_{Li}}{C_{Li}}}{\dfrac{dw}{w} - \dfrac{dr}{r}} \qquad (i = 1, 2, 3)$$

From our assumption of quasi concavity of the production functions, $\sigma_i > 0$.

A. *Geometrical Derivation of the Transformation Curve*

The principal issues in this section concern the shape and the slope of the transformation curve, defined as the locus of all efficient combinations of the final commodities satisfying two conditions, namely, i) the full employment of the primary factors, and ii) $D_3 = X_3$. Since there are actually three commodities produced and three factors (two primary inputs and one pure intermediate good), the derivation of the economy's transformation curve in our model is more complicated than in the traditional two-commodity, two-primary-factor model. Our first task will thus be that of deriving the locus of efficient outputs.

Here a distinction must be drawn between the gross and the net transformation curves. For expository purposes, assume that all three goods are produced only with the help of the two primary factors. We can then obtain a smooth, concave to the origin, "ruled" production possibility surface such as ABC in Figure 1. In each coordinate plane, the set of possible production points constitutes a gross transformation curve for the two relevant commodities.[2] For example, AB is the locus of all feasible combinations of X_1 and X_2 if no X_3 is produced, so that AB may be regarded as the gross transformation curve in our model.

Assume now that the production of X_2 requires the use of the intermediate good, X_3, in the proportion a_2 (i.e., $(X_{32}/X_2 = a_2)$. The maximum net output of X_2 must obviously be smaller than OB. Draw the line OD in the X_2X_3 plane such that the

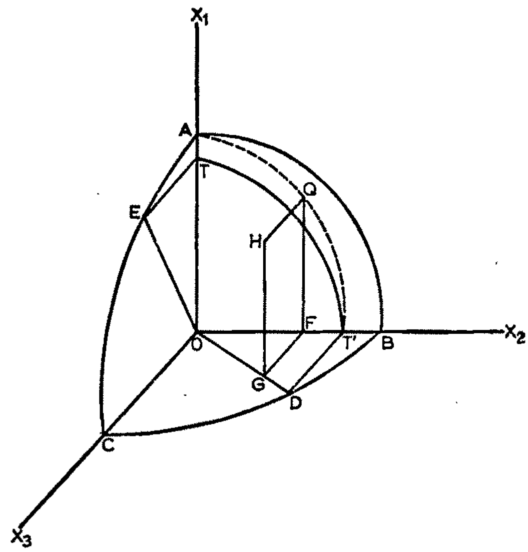


FIGURE 1

slope of OD with respect to the X_2 axis equals the given a_2; that is, $(DT'/OT') = a_2$. The maximum net output of X_2 is then seen to be OT', which utilizes DT' of the intermediate good, X_3, where D, being a point on the production surface, is both feasible and efficient.

Assume now that the net output of X_2 is actually OF, where OF is any arbitrary level of X_2 less than OT'. The amount of X_3 needed to produce this output level is given by FG. Since G does not lie on the production surface, the unemployed primary factors can be used in the production of the first commodity in an amount given by the vertical distance GH, from G to the production surface. By drawing HQ parallel to FG we obtain point Q in the X_1X_2 plane, which shows that with full employment of the two primary factors and a net output of X_2 equal to OF, the maximum net output of X_1 is $FQ(=GH)$. By repeating this construction for all possible net outputs of X_2 we may derive AT', which is the net transformation curve drawn under the assumption that $a_2 > 0$ and $a_1 = 0$.

The next step is to allow for the use of the intermediate input in the production of

[2] For further details on the derivation and construction of the production possibility surface in a model with three goods and two primary factors, see Melvin (1968).

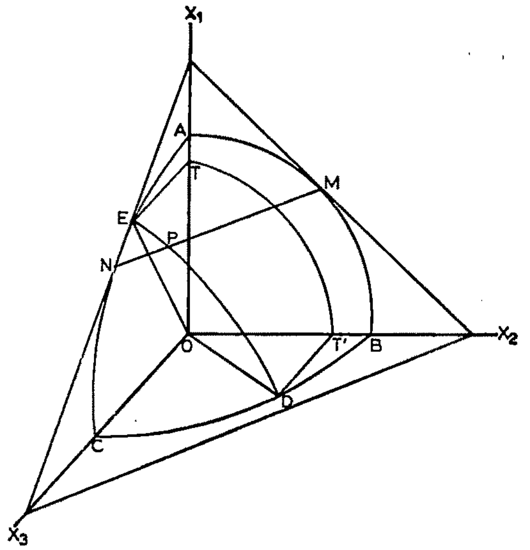


FIGURE 2

X_1 also. Let the slope of the line OE, with respect to the X_1 axis in the X_1X_3 plane, equal the given level of a_1 (i.e.,$ET/OT = a_1$); OT is then the maximum net output of X_1. By using the same procedure through which AT' was derived, given AB as the gross transformation curve, we can now obtain TT', given AT'. The curve TT' is then the net transformation curve for the economy.

Before proceeding to a rigorous analysis of the slope and the shape of this production possibility locus, a geometrical proof of the convexity of the production set will be given with the use of Figure 2, which reproduces the main features of Figure 1. The satisfaction of the full employment condition requires all efficient production points to be on the ABC surface. On the other hand, the condition that the demand for the intermediate product equals its supply requires that output combinations satisfy the equation

$$a_1X_1 + a_2X_2 = X_3 \tag{3}$$

Thus, any feasible and efficient point must lie on the interesection of the production surface and the plane given by equation (3). Melvin (1968) has shown that the production set in a three-good, two-primary-factor model is convex when production functions exhibit constant returns to scale. The intersection of such a set and a plane must therefore itself be convex—this intersection being the area OED in Figure 2. Thus, the curve ED is concave to the origin, and its projection on the X_1X_2 plane, the curve TT', must also be concave to the origin.

Figure 2 also illustrates the difference between the three-final-commodity model and the one described here with respect to the indeterminacy of output levels for given commodity prices. As Melvin (1968) has shown, any price plane will be tangent to the ABC surface along a line, such as MN. In our model, however, only point P along MN is relevant, because P is the only output combination satisfying (3). The latter constraint constitutes, in effect, a "demand" condition which solves the indeterminacy problem.

B. *The Slope and the Shape of the Transformation Curve*

The intuitive explanation concerning the shape of the transformation curve and the indeterminacy problem presented above may now be supplemented by a rigorous proof. However, the first step consists in obtaining an expression for the slope of the transformation curve. It has been shown by Vanek that the slope of the net transformation curve equals the negative of the commodity-price ratio even if the final commodities also serve as intermediate inputs. We will now show that this result holds in our model where the material input is not identical with the final goods.

Since national income may be measured as the value of final outputs or the sum of payments to primary factors, we can write

$$Y = X_1 + p_2X_2 = w\overline{L} + r\overline{K} \tag{8}$$

Differentiating this expression, keeping in mind that factor supplies are given, we obtain

$$(9)\quad dX_1 + p_2 dX_2 + X_2 dp_2 = \bar{L}dw + \bar{K}dr$$

Substituting for $\bar{L}$ and $\bar{K}$ from (1) and (2), we have

$$\begin{aligned}(10)\quad &\bar{L}dw+\bar{K}dr\\ &=(C_{L1}X_1+C_{L2}X_2+C_{L3}X_3)dw\\ &\qquad +(C_{K1}X_1+C_{K2}X_2+C_{K3}X_3)dr\\ &=(C_{L1}dw+C_{K1}dr)X_1\\ &\qquad +(C_{L2}dw+C_{K2}dr)X_2\\ &\qquad +(C_{L3}dw+C_{K3}dr)X_3\end{aligned}$$

Totally differentiating the price equations (4)–(6), and remembering that $dp_1=0$, we obtain

$$\begin{aligned}(11)\quad & C_{L1}dw + C_{K1}dr + a_1 dp_3 = \quad - (wdC_{L1} + rdC_{K1})\\ (12)\quad & C_{L2}dw + C_{K2}dr + a_2 dp_3 = dp_2 - (wdC_{L2} + rdC_{K2})\\ (13)\quad & C_{L3}dw + C_{K3}dr = dp_3 - (wdC_{L3} + rdC_{K3})\end{aligned}$$

It must be noted, however, that the equilibrium minimum-unit-cost condition of equality between the marginal rate of substitution among two factors and the ratio of these factors' prices implies that

$$(14)\quad wdC_{Li} + rdC_{Ki} = 0 \qquad (i = 1, 2, 3)$$

Using (14) and substituting (11)–(13) into (10) then yields

$$\begin{aligned}(15)\quad &\bar{L}dw+\bar{K}dr\\ &=-a_1X_1dp_3+X_2(dp_2-a_2dp_3)+X_3dp_3\\ &=-(a_1X_1+a_2X_2)dp_3+X_2dp_2+X_3dp_3\\ &=X_2dp_2\end{aligned}$$

From (9) and (15), we now obtain

$$(16)\quad dX_1 + p_2 dX_2 = 0$$

or

$$(16^*)\quad \frac{dX_1}{dX_2} = - p_2$$

which establishes the traditional result that the marginal rate of transformation between the final commodities equals the ratio of their prices (since $p_1=1$) in equilibrium.

Given the relationship expressed in (16*), the shape of the transformation curve can now be derived from the response of outputs to changes in their prices. In particular, if the output of a commodity rises in response to an increase in its relative price, and conversely, the net transformation curve is concave to the origin. Otherwise, it is convex to the origin.[3] Therefore, our task here is to find out the effect of a change in p_2 on the output of X_1 and X_2.

Solving (11)–(13) simultaneously for dw and dr, we obtain

$$(17)\quad dw = \frac{(C_{K1} + a_1 C_{K3})}{A}\, dp_2$$

$$(18)\quad dr = \frac{-(C_{L1} + a_1 C_{L3})}{A}\, dp_2$$

where

$$\begin{aligned}A &= C_{L1}C_{L2}(k_1 - k_2) + a_1 C_{L2}C_{L3}(k_3 - k_2)\\ &\qquad + a_2 C_{L1}C_{L3}(k_1 - k_3)\end{aligned}$$

and $k_i=K_i/L_i=C_{Ki}/C_{Li}$ is the capital-labor ratio in the ith industry.

We may now differentiate the full employment equations (1)–(3), keeping in mind that $\bar{L}$ and $\bar{K}$ are given, to obtain

$$(19)\quad C_{L1}dX_1 + C_{L2}dX_2 + C_{L3}dX_3 =$$

[3] However, if $dX_1/dX_2 \neq -p_2$ the shape of the transformation curve is not dependent exclusively on the response of outputs to changes in commodity prices. For example, Bhagwati and T. Srinivasan, Kemp and H. Herberg, and Ronald Jones (1971) have shown that in the presence of an interindustry wage differential, the transformation curve may be locally or globally convex to the origin even if the output responses to price changes are normal, and vice versa.

$$= -(X_1 dC_{L1} + X_2 dC_{L2} + X_3 dC_{L3})$$
$$= R_L\left(\frac{dw}{w} - \frac{dr}{r}\right)$$

$$(20)\quad C_{K1}dX_1 + C_{K2}dX_2 + C_{K3}dX_3$$
$$= -(X_1 dC_{K1} + X_2 dC_{K2} + X_3 dC_{K3})$$
$$= -R_K\left(\frac{dw}{w} - \frac{dr}{r}\right)$$

and

$$(21)\quad a_1 dX_1 + a_2 dX_2 - dX_3 = 0$$

where R_L and R_K are derived in the Appendix and are both positive. From (17) and (18), it is readily shown that

$$(22)\quad \left(\frac{dw}{w} - \frac{dr}{r}\right)$$
$$= \frac{(C_{K1}+a_1C_{K3})r+(C_{L1}+a_1C_{L3})w}{wr\cdot A}\,dp_2$$
$$= \frac{dp_2}{wr\cdot A}$$

since $C_{L3}w+C_{K3}r=p_3$ from (6) and $p_1=1$ by assumption.

We can now solve (19)–(21) simultaneously for dX_1 and dX_2, making use of (22), to derive:

$$(23)\quad \frac{dX_1}{dp_2} = \frac{-1}{wr\cdot A^2}\,[(C_{K2} + a_2C_{K3})R_L + (C_{L2} + a_2C_{L3})R_K]$$

$$(24)\quad \frac{dX_2}{dp_2} = \frac{1}{wr\cdot A^2}\,[(C_{K1} + a_1C_{K3})R_L + (C_{L1} + a_1C_{L3})R_K]$$

If the output responses to changes in commodity prices are to be "normal,"[4] then $dX_1/dp_2<0$ and $dX_2/dp_2>0$. From (23) and (24), it is clear that these inequalities are indeed satisfied in our model and, therefore, the output responses to a change in relative prices are normal, or equivalently, the net transformation curve is strictly concave to the origin. Witness further that (23) and (24) exhibit a monotonic relationship between the relative prices of the two final commodities and their outputs. This shows that there is only one configuration of commodity outputs consistent with one set of commodity prices. In other words, there is no indeterminacy problem present in our model.[5]

II. The Rybczynski Theorem

According to Rybczynski an increase in the supply of a primary factor of production promotes, at constant commodity prices, an expansion of the output of the commodity which uses the expanding factor relatively intensively at the expense of the output of the other commodity. We now explore the consequences of a rise in the supply of a factor for the output of the final goods, X_1 and X_2, in our model with intermediate products.[6] Now, if prices of final goods are kept constant ($dp_2=0$), the price of the intermediate good as well as the wage and rental rates remain constant,[7] so that the production coefficients will also remain unchanged. Differentiating (1)–(3) with respect to $\overline{K}$ and $\overline{L}$ then yields:

[4] By normal output response we mean that the supply curve for a final commodity has its usual positive slope.

[5] The need for an elaborate proof of the absence of an indeterminacy problem in our model arose from several questions raised by readers of an earlier draft of this paper. It may be emphasized here that even though our model is characterized by an inequality between the number of commodities and primary factors—an inequality whose implications for the indeterminacy problem have been thoroughly analyzed by Melvin (1968)—yet the model is completely determinate because the output of the third commodity, which serves as an intermediate input, is determined by the output of the two final commodities. Thus, in a sense, our model comprises three commodities and three inputs.

[6] For other extensions of the Rybczynski theorem, see Batra (1970), Kemp (1969), Jones (1968), and Casas (1972b).

[7] For a rigorous proof of the relationship between the prices of final goods and that of the intermediate product, see Section IV on the pattern of trade.

$$(25)\qquad \frac{\partial X_1}{\partial K} = \frac{(C_{L2} + a_2C_{L3})}{A}$$

$$(26)\qquad \frac{\partial X_2}{\partial K} = \frac{-(C_{L1} + a_1C_{L3})}{A}$$

$$(27)\qquad \frac{\partial X_3}{\partial K} = \frac{(a_1C_{L2} - a_2C_{L1})}{A}$$

$$(28)\qquad \frac{\partial X_1}{\partial L} = \frac{-(C_{K2} + a_2C_{K3})}{A}$$

$$(29)\qquad \frac{\partial X_2}{\partial L} = \frac{(C_{K1} + a_1C_{K3})}{A}$$

$$(30)\qquad \frac{\partial X_3}{\partial L} = \frac{(a_2C_{K1} - a_1C_{K2})}{A}$$

where, as before,

$$A = C_{L1}C_{L2}(k_1 - k_2) + a_1C_{L2}C_{L3}(k_3 - k_2) + a_2C_{L1}C_{L3}(k_1 - k_3)$$

It may be observed that the numerators of (25) and (29) are positive, and those of (26) and (28) are negative, so that the signs of $\partial X_i/\partial K$ and $\partial X_i/\partial L$ $(i=1, 2)$ depend on the sign of A. Examination of the latter reveals that if k_3 lies between k_1 and k_2, then A has a definite sign. In particular, if $k_1 \gtrless k_3 \gtrless k_2$. $A \gtrless 0$, so that $\partial X_1/\partial K \gtrless 0$, $\partial X_2/\partial K \lessgtr 0$, $\partial X_1/\partial L \lessgtr 0$, and $\partial X_2/\partial L \gtrless 0$.

THEOREM 1: *Thus, we may conclude that if the capital-labor ratio of the intermediate product lies in between the other two capital-labor ratios, the Rybczynski theorem holds: An increase in the stock of capital (labor), at constant prices, raises (lowers) the output of X_1 and lowers (raises) the output of X_2 if X_1 is capital-intensive relative to X_2, and vice versa.*

The effect of factor growth on the output of X_3 depends not only on the sign of A, but also on $a_i/C_{Li} = X_{3i}/L_i$ (in the case of capital accumulation) and $a_i/C_{Ki} = X_{3i}/K_i$ (in the case of growth of the labor force). Let us consider the case of an increase in the stock of capital. Then, if $A > 0$—so that the output of X_1 rises—$\partial X_3/\partial K \gtrless 0$ if $a_1/C_{L1} \gtrless a_2/C_{L2}$, and conversely if $A < 0$ and the output of X_2, instead of X_1, rises.

If k_3 does not lie between k_1 and k_2, then the Rybczynski theorem may not hold. If $a_1/C_{L1} = a_2/C_{L2}$, then $A = C_{L2}(C_{L1} + a_1C_{L3})(k_1 - k_1)$ and the theorem always holds. But this is a trivial case because, here, the output of X_3 is constant. The difficulty arises when $a_1/C_{L1} \neq a_2/C_{L2}$ and k_3 lies outside the confines of k_1 and k_2. There are two possibilities: (i) $k_1 \gtrless k_2 \gtrless k_3$ and (ii) $k_2 \gtrless k_1 \gtrless k_3$. It turns out that in case (i) the Rybczynski theorem holds if $a_2/C_{L2} > a_1/C_{L1}$, whereas in case (ii) its validity requires that $a_1/C_{L1} < a_2/C_{L2}$. Take, for example, the case where $k_1 > k_2 > k_3$. Here $(k_1 - k_2) > 0$ and if $a_2/C_{L2} > a_1/C_{L1}$, we can see that $a_1C_{L2}(k_2 - k_3) < a_2C_{L1}(k_1 - k_3)$ because $(k_1 - k_3) > (k_2 - k_3)$. Hence, $A > 0$, $\partial X_1/\partial K > 0$ and $\partial X_2/\partial K < 0$, $\partial X_1/\partial L < 0$ and $\partial X_2/\partial L > 0$. Therefore, the Rybczynski theorem is valid. This leads to the following conclusion:

THEOREM 2: *If a commodity is relatively intensive in the use of the intermediate good, and if the capital-labor ratio of that commodity lies between the capital-labor ratio of the intermediate product and that of the other final commodity, the Rybczynski theorem holds.*

In Theorem 2, intensity in the use of the intermediate good is measured by $a_i/C_{Li} = X_{3i}/L_i (i=1, 2)$.[8] It is evident that if the

[8] Alternatively, intensity in the use of the intermediate good can be expressed in terms of the intermediate input/capital ratio, $a_i/C_{Ki} = X_{3i}/K_i$. Thus, A may be rewritten as

$$A = C_{K1}C_{K2}\left(\frac{1}{k_2} - \frac{1}{k_1}\right) + a_1C_{K2}C_{K3}\left(\frac{1}{k_2} - \frac{1}{k_3}\right) + a_2C_{K1}C_{K3}\left(\frac{1}{k_3} - \frac{1}{k_1}\right)$$

If $k_1 > k_2 > k_3$, for example, a sufficient condition for $A > 0$ is that $a_2/C_{K2} > a_1/C_{K1}$. Hence, Theorem 2 remains unaffected by the change in the definition of intensity in the use of the intermediate good.

conditions implied in Theorem 2 are not satisfied, then the Rybczynski theorem may not hold.

What is the economic explanation of the fact that, unlike in Vanek's interindustry flows framework, the Rybczynski theorem may not be valid in our model which allows for the presence of pure intermediate goods? The answer to this question is facilitated by introducing a distinction between the net capital-labor ratio and the gross capital-labor ratios in the final products. The net or apparent capital-labor ratio of the ith final commodity is given by $k_i = K_i/L_i = C_{Ki}/C_{Li}$, whereas the gross or total capital-labor ratio, k_i^*, equals the ratio between the amounts of capital and labor utilized directly and indirectly by the ith final good. For example, the gross amount of capital employed by the ith final commodity, K_i^*, equals K_i plus the amount of capital employed in the intermediate good (i.e., K_3) multiplied by the proportion of the intermediate product used in the production of the ith final good, that is,

$$K_i^* = K_i + (X_{3i}/X_3)K_3$$

Similarly, the gross amount of labor utilized by the ith final good is given by

$$L_i^* = L_i + (X_{3i}/X_3)L_3$$

Hence,

$$k_i^* = \frac{K_i + (X_{3i}/X_3)K_3}{L_i + (X_{3i}/X_3)L_3} = \frac{k_iC_{Li} + a_ik_3C_{L3}}{C_{Li} + a_iC_{L3}} \qquad (i = 1, 2)$$

Thus, if the first commodity is apparently capital-intensive relative to the second commodity (that is, if $k_1 > k_2$), then it will be truly capital intensive (that is, $k_1^* > k_2^*$) only if

$$\frac{k_1C_{L1} + a_1k_3C_{L3}}{C_{L1} + a_1C_{L3}} > \frac{k_2C_{L2} + a_2k_3C_{L3}}{C_{L2} + a_2C_{L3}},$$

or

$$C_{L1}C_{L2}(k_1 - k_2) + a_1C_{L2}C_{L3}(k_3 - k_2) + a_2C_{L1}C_{L3}(k_1 - k_3) = A > 0$$

In other words, conditions which are sufficient to ensure that $A > 0$ will also ensure that the first commodity is relatively capital-intensive both in net and gross terms. We have previously established that $A > 0$ if (i) $k_1 > k_3 > k_2$, or (ii) $k_1 > k_2 > k_3$ *and* $a_2/C_{L2} \geq a_1/C_{L1}$. From this discussion the following general result is immediate:

The factor-intensity rankings of the final products in the net and gross sense are identical if (i) k_3 lies between k_1 and k_2, or (ii) the commodity whose capital-labor ratio lies between the capital-labor ratios of the intermediate product and the other final product is at least as intensive in the use of the intermediate product as the other final commodity. Any one of these conditions is also sufficient to give rise to an unambiguous sign of A.

If condition (ii) above is not satisfied there arises the possibility of a conflict between the factor-intensity rankings in the net and the gross sense; that is to say, a commodity may be relatively capital-intensive in the net sense, but relatively labor-intensive in the gross sense, and vice versa. This possibility of a conflict between the net and the gross factor-intensity rankings of the final products constitutes the basic difference between our model with pure intermediate products and Vanek's interindustry flows model. As shown originally by Vanek and recently by Casas (1972a) the gross and the net capital-labor ratios in the final products always enjoy the same ranking in a model where interindustry flows are present. In our model with pure intermediate products, however, the net or the apparent factor-intensity ranking of a final commodity may not reflect its true or gross factor-intensity ranking.

It must be clear by now that the Rybczynski theorem will be invalid only if the net and the gross factor-intensity rankings run counter to each other; in other words, a rise in the supply of, say, capital alone may lower the output of an apparently capital-intensive commodity at constant prices, if in the gross sense this commodity is relatively labor-intensive. Needless to say, the Rybczynski theorem is always valid if factor-intensity rankings are defined in the gross sense.

III. The Stolper-Samuelson Theorem

According to the Stolper-Samuelson theorem, a rise in the relative price of a commodity raises the real reward of the primary factor employed more intensively by it and lowers the real reward of the other primary factor, and conversely. We will now examine the implications of a change in the price ratio between the final products for factor rewards in our model incorporating intermediate goods.[9]

Let us assume that the relative price of the second commodity increases, i.e., $dp_2>0$; from (4)–(6) and (17)–(18) the change in real factor rewards in terms of the higher priced commodity is given by:

$$\frac{d(w/p_2)}{dp_2} = \frac{(C_{K2} + a_2C_{K3})}{p_2^2 \cdot A} \tag{31}$$

$$\frac{d(r/p_2)}{dp_2} = \frac{-(C_{L2} + a_2C_{L3})}{p_2^2 \cdot A} \tag{32}$$

For the Stolper-Samuelson theorem to hold, $d(w/p_2)/dp_2>0$ and $d(r/p_2)/dp_2<0$ if $k_1>k_2$, and vice versa. The numerator of (31) is positive and that of (32) is negative, so that the signs of $d(w/p_2)/dp_2$ and $d(r/p_2)/dp_2$ depend on the sign of A. In other words, the Stolper-Samuelson theorem holds if A has the same sign as (k_1-k_2) or, what is the same thing, if the net and gross factor-intensity rankings are identical; otherwise, the Stolper-Samuelson theorem may not hold.

Another result worth investigating is the so-called fundamental duality theorem. In terms of a two-commodity, two-factor model, Jones (1965) has shown that there exists a dual relationship between the Rybczynski theorem and the Stolper-Samuelson theorem in the sense that the effect of a change in factor endowments on commodity outputs at constant prices is identical to the effect of a change in commodity prices on nominal factor rewards at constant factor endowments. It is interesting to see that this duality remains unscathed in our generalized model where pure intermediate goods are allowed. This may be demonstrated simply by comparing (18) and (27) to find that

$$\frac{\partial X_2}{\partial K} = \frac{dr}{dp_2} = \frac{-(C_{L1} + a_1C_{L3})}{A} \tag{33}$$

IV. The Pattern of Trade

It is fashionable to attempt to explain the pattern of trade between two countries in terms of the Heckscher-Ohlin theorem: a country exports the commodity which uses intensively its relatively abundant factor, and imports the commodity which is intensive in the use of its relatively scarce factor. Traditionally, the logical validity of this theorem has been shown to depend crucially upon the validity of the Stolper-Samuelson and the Rybczynski theorems. Unfortunately, in our model where three goods are produced, the pattern of trade turns out to be indeterminate if all three goods are to be traded, even if the Rybczynski and the Stolper-Samuelson theorems hold unambiguously. To get an immediate idea of this problem, let us revert to the price equations (4)–(6). If all three goods are traded, p_1, p_2, and p_3 are all determined exogenously. The three equations (4)–(6) then contain three pa-

[9] For other extensions of the Stolper-Samuelson theorem, see Nobuo Minabe, Batra (1968), Kemp (1969) and Jones (1968).

rameters, p_1, p_2, and p_3, to solve for two unknowns w and r in terms of the input-output coefficients, so that the model becomes indeterminate.

One solution to this problem is to assume that world prices are such that the production of one commodity is unprofitable, so that it is no longer produced. But this is not attractive, for then we fall into the domain of the traditional two-good model. Another solution lies in the stipulation that one of the three goods is nontraded. For example, suppose the intermediate good is not traded at all. Then the three equations (4)–(6) containing two parameters p_1 and p_2 can be solved to obtain the three unknowns w, r, and p_3. The number of equations matches the number of unknowns, and the indeterminacy problem disappears.

A closely related, though not completely identical, question concerns the determination of the pattern of trade when all three goods are to enter trade. Suppose initially no trade in X_3 is allowed. If we assume that international production functions are identical, then the Heckscher-Ohlin theorem follows directly from the validity of the Stolper-Samuelson theorem. If, in addition, we assume that consumption patterns are also identical internationally, then the Heckscher-Ohlin theorem derives from the validity of the Rybczynski theorem. If all three goods continue to be produced in free trade equilibrium, factor prices are equated between countries; X_3 also commands the same price between countries, and there is no incentive to trade X_3 once the prohibition to such trade is lifted.

Nonetheless, trade *could* take place in either direction. World outputs and world demand for final products are unaffected, because the free trade equilibrium gives rise to a given set of world prices. But each country's transformation curve shifts—outwards for the country importing X_3 and inwards for the other. Suppose there are two countries, the home country and the foreign country, and suppose further that the former is relatively capital-abundant and that $k_1 > k_3 > k_2$. If X_3 is exported by the home country, at free trade prices the home production of both X_1 and X_2 declines (and rises abroad). Thus the home country might eventually import *both* X_1 and X_2 in exchange for X_3, although the Heckscher-Ohlin theory would require the home country to *export* X_1. Alternatively, if the home country imports X_3, it might (i) be an exporter of X_1 and an importer of X_2, or (ii) export both X_1 and X_2 and import X_3. Thus without further restrictions it is impossible to pick out *any* commodity and state whether it is imported or exported by the home country.

The solution that we wish to propose must have become obvious by now. If one of the three goods is nontraded, not only do we solve the indeterminacy problem in the model, but we can also salvage the Heckscher-Ohlin theory. Note that our solution does not require that the nontraded good be the intermediate good only, although, as it turns out, if the nontraded good is one of the final goods, the validity of the Heckscher-Ohlin theorem is not all that obvious.

Let us define the home country to be capital-abundant relative to the foreign country if

$$\omega_h > \omega_f$$

where $\omega = w/r$ is the wage-rental ratio and the subscripts h and f refer, respectively, to the home and foreign countries. If the intermediate good is not traded, all that we need to demonstrate in order to prove the Heckscher-Ohlin theorem is that $(d\omega/dp_2)$ has an unambiguous sign. From (17) and (18)

$$\text{(34)} \quad \frac{d\omega}{dp_2} = \left(r\frac{dw}{dp} - w\frac{dr}{dp}\right) \bigg/ r^2 = \frac{1}{r^2 A}$$

From (34) it is clear that the relationship between ω and p_2 is unique only if A has an unambiguous sign. If $A > 0$, then with $\omega_h > \omega_f$, it is clear that

$$p_{2h} > p_{2f}$$

where p_{2j} is the autarky relative price of the second good in terms of the first in the jth country ($j = h, f$). The home country will then export X_1 and import X_2. On the other hand, if $A < 0$, then with $\omega_h > \omega_f$,

$$p_{2h} < p_{2f},$$

so that the home country will export X_2 and import X_1. In general, with the intermediate good nontraded, the Heckscher-Ohlin theorem will hold if sign $(A) = $ sign $(k_1 - k_2)$ and we already know that a sufficient condition for this is that k_3 lies between k_1 and k_2.

Let us now consider the other case where one of the final goods is the nontraded good. Let X_1 be such a good. For the demonstration of the Heckscher-Ohlin theorem, we now need an expression relating p_3/p_2 with ω. From (13), (17), and (18) it is readily shown that

$$\text{(35)} \qquad \frac{dp_3}{dp_2} = \frac{C_{L1}C_{L3}(k_1 - k_3)}{A}$$

Dividing (35) by (34), we have

$$\text{(36)} \qquad \frac{dp_3}{d\omega} = r^2 C_{L1}C_{L3}(k_1 - k_3)$$

and from (34) and (36)

$$\text{(37)} \qquad \frac{d(p_3/p_2)}{d\omega} = \frac{1}{p_2^2}\left[p_2\frac{dp_3}{d\omega} - p_3\frac{dp_2}{d\omega}\right]$$
$$= \frac{r^2}{p_2^2}[(p_2 - a_2p_3)C_{L1}C_{L3}(k_1 - k_3) + p_3C_{L1}C_{L2}(k_2 - k_1) + a_1p_3C_{L2}C_{L3}(k_2 - k_3)]$$

If $\omega_h > \omega_f$ and $k_2 > k_3$, then from the Heckscher-Ohlin theorem we would expect the home country to export X_2 and import X_3. For this pattern of trade to take place, we require that

$$\frac{d(p_3/p_2)}{d\omega} > 0$$

so that $(p_3/p_2)_h > (p_3/p_2)_f$ when $\omega_h > \omega_f$. However, $k_2 > k_3$ alone is not sufficient to ensure this. A sufficient condition for $d(p_3/p_2)/d\omega > 0$ is that $k_2 > k_1 > k_3$, that is, X_2 is the most capital-intensive and X_3 the least capital-intensive of all three goods. Conversely, if $k_2 < k_3$, the validity of the Heckscher-Ohlin theory requires that $d(p_3/p_2)/d\omega < 0$, and this is sufficiently ensured if $k_2 < k_1 < k_3$.

A definite pattern emerges from our discussion concerning the validity of the Heckscher-Ohlin theorem when one of the goods is specified as being the nontraded good. This is summarized in the following theorem:

THEOREM 3: *If one of the goods is specified as being nontraded (whether it is a final or an intermediate good), the international exchange of the other two will follow the Heckscher-Ohlin dictum if the capital-labor ratio of the nontraded good lies between the capital-labor ratios of the other two goods.*

If this theorem is not satisfied, the Heckscher-Ohlin theorem may not hold even if one of the three goods is specified as the nontraded good.

The Nature of the Free Trade Equilibrium

Given that one commodity is nontraded, what is the nature of equilibrium in free trade? If the nontraded good is the intermediate product, the answer is very simple and conforms to the traditional graphical explanations, where the world prices in conjunction with the transformation curve determine the free trade production point, and where the consumption point is deter-

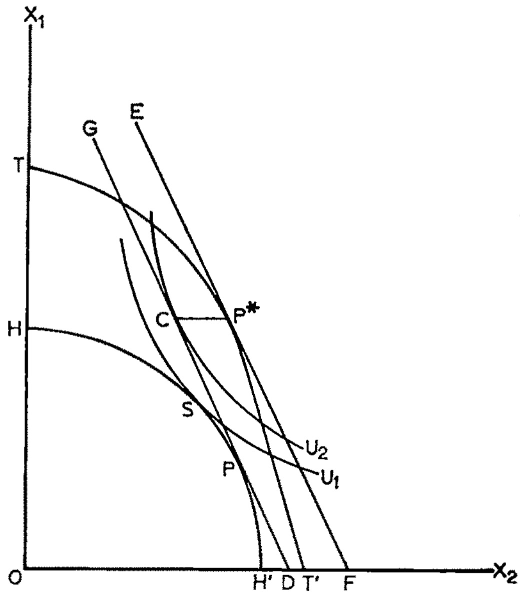


FIGURE 3

mined by the demand conditions. The problem is not so simple when the nontraded good is one of the final commodities. The free trade consumption point relating to final goods is, as before, determined by the demand conditions, but the production point is determined not only by world prices but also by the volume of trade.

As an illustration, consider Figure 3 which depicts the case of the home country exporting X_2 and importing X_3 while treating X_1 as the nontraded good. As previously seen, such a pattern would arise if $\omega_h > \omega_f$ under autarky and $k_2 > k_1 > k_3$. In Figure 3, HH' is the home country's net transformation curve and S is the point of self-sufficiency equilibrium which is determined by the tangency of the social indifference curve U_1 with HH'. Since the home country exports X_2 and imports X_3, the switch from autarky to free trade causes a rise in p_2 and a decline in p_3, with all prices expressed in terms of the nontraded good. For the time being assume that no X_3 is imported. Then the rise in p_2 as a result of the introduction of trade shifts production from S to P where GD, the price line prevailing in the free trade equilibrium, is tangent to the net transformation curve, and consumption moves from S to C which lies on a higher indifference curve U_2. The final equilibrium production point is not P, however, for the imports of X_3 will shift the transformation curve outwards, the extent of this shift depending upon the volume of X_3 imported. Such a transformation curve is given by TT' and the final production point is given by P^* where the price line EF, parallel to GD, touches the new transformation curve.

Several aspects deserve further exposition. First, the free trade equilibrium production point P^* is horizontally aligned with the consumption point C, showing that the supply of and the demand for X_1 are equal. Second, the output of X_2 corresponding to P^* may be smaller or larger than that corresponding to P. This follows from the fact that the import of X_3 is equivalent to the injection of capital and labor in the home country in the proportion $C_{K3}/C_{L3} = k_3$. However, with $k_2 > k_1 > k_3$, the imports of X_3 will result in a decline of the output of X_2 and a rise in the output of X_1 at the free trade prices if X_2 is capital-intensive in gross terms,[10] and vice versa. For this reason, if we assume that sign $(A) =$ sign $(k_1 - k_2)$, it follows that the shift in the transformation curve

[10] If we rewrite equation (3) as

$$X_3 = a_1 X_1 + a_2 X_2 - E_3$$

where E_3 denotes the imports of the intermediate good, then at constant prices and with fixed supplies of primary factors it is readily shown that

$$\frac{\partial X_1}{\partial E_3} = -\frac{C_{L2} C_{L3}(k_2 - k_3)}{A}$$

and

$$\frac{\partial X_2}{\partial E_3} = \frac{C_{L1} C_{L3}(k_1 - k_3)}{A}$$

Thus, $\partial X\ /\partial E_3 \gtrless 0$ and $\partial X_2/\partial E_3 \lesseqgtr 0$ if $A \lesseqgtr 0$.

from HH' to TT' will be biased towards X_1. Third, the length CP^* measures the value of the imports of X_3 in terms of the second commodity and, in equilibrium, this equals the value of the exports of X_2. This follows from the fact that the change in the value of home production, equal to CP^* in Figure 3, must be equal to the amount of the increase in the availability of the intermediate input multiplied by its price. Expressed in terms of the second commodity, this equals p_3E_3/p_2, where E_3 is the amount of X_3 imported, that is,[11]

$$CP^* = p_3E_3/p_2$$

But p_3E_3/p_2 is nothing but the value of the imports of E_3 expressed in terms of the second commodity. With balanced trade, this must equal the export of X_2 which, in Figure 3, is also given by CP^*. In other words, the location of the final production point P^* satisfies the balance of trade equilibrium condition, namely, $p_3E_3=p_2E_2$, where E_2 is the amount of X_2 exported. In toto, Figure 3 shows that as a result of the switch from autarky to free trade, production moves from S to P^*, consumption moves from S to C, and welfare improves from U_1 to U_2.

An analogous but opposite procedure can be followed for the case where the home country is an exporter of X_3 and an importer of X_2. Here, $k_2<k_1<k_3$ with $\omega_h>\omega_f$ under autarky. This case is pictured in Figure 4 where the switch from autarky to free trade takes production from S to P^*, consumption from S to C, and welfare from U_1 to U_2. The distance CP^* equals the imports of X_2 as well as the value of the exports of X_3 expressed in terms of the second commodity, with the demand for X_1 equal to the domestic production of X_1.

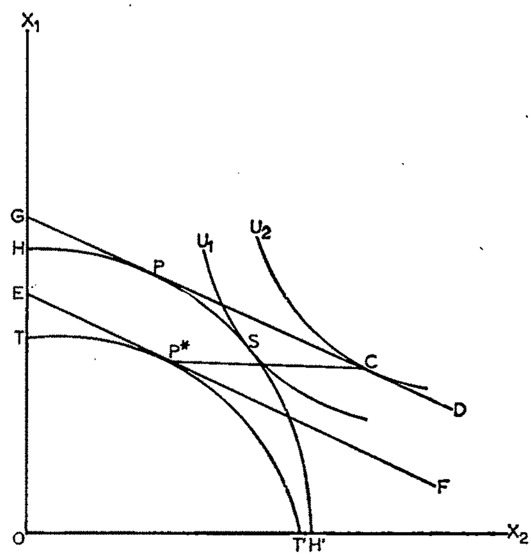


FIGURE 4

[11] At constant prices, the increase in the total value of domestic output resulting from the importation of E_3 units of the intermediate good is given by

$$\left(\frac{\partial X_1}{\partial E_3} + p_2\frac{\partial X_2}{\partial E_3}\right) E_3$$

This is easily shown to equal p_3E_3 in terms of the first commodity and, hence, p_3E_3/p_2 in terms of the second good.

APPENDIX

From the definition of σ_i (equation (7)), the minimum-unit-cost conditions (equation (14)), and the price equations (4)–(6), the following results may be derived:

$$\frac{dC_{K1}}{C_{K1}} = \sigma_1\left(\frac{dw}{w} - \frac{dr}{r}\right) + \frac{dC_{L1}}{C_{L1}}$$

$$= \sigma_1\left(\frac{dw}{w} - \frac{dr}{r}\right) - \frac{r}{w}\,\frac{dC_{K1}}{C_{L1}}$$

Hence,

$$\frac{dC_{K1}}{wC_{L1}C_{K1}}(wC_{L1} + rC_{K1}) = \sigma_1\left(\frac{dw}{w} - \frac{dr}{r}\right)$$

or

$$dC_{K1} = \frac{wC_{L1}C_{K1}\sigma_1}{(1 - a_1p_3)}\left(\frac{dw}{w} - \frac{dr}{r}\right)$$

In a similar way, it is readily shown that

$$dC_{L1} = \frac{-rC_{L1}C_{K1}\sigma_1}{(1 - a_1p_3)}\left(\frac{dw}{w} - \frac{dr}{r}\right)$$

$$dC_{K2} = \frac{wC_{L2}C_{K2}\sigma_2}{(p_2 - a_2p_3)}\left(\frac{dw}{w} - \frac{dr}{r}\right)$$

$$dC_{L2} = \frac{-rC_{L2}C_{K2}\sigma_2}{(p_2 - a_2p_3)}\left(\frac{dw}{w} - \frac{dr}{r}\right)$$

$$dC_{K3} = wC_{L3}C_{K3}\sigma_3\left(\frac{dw}{w} - \frac{dr}{r}\right)$$

$$dC_{L3} = -rC_{L3}C_{K3}\sigma_3\left(\frac{dw}{w} - \frac{dr}{r}\right)$$

Using these relationships, we may now derive:

$$= \frac{-(X_1dC_{L1} + X_2dC_{L2} + X_3dC_{L3})}{\left(\frac{dw}{w} - \frac{dr}{r}\right)}$$

$$= r\left[\frac{X_1C_{L1}C_{K1}\sigma_1}{(1 - a_1p_3)} + \frac{X_2C_{L2}C_{K2}\sigma_2}{(p_2 - a_2p_3)} + X_3C_{L3}C_{K3}\sigma_3\right] > 0$$

and

$$R_K = \frac{(X_1dC_{K1} + X_2dC_{K2} + X_3dC_{K3})}{\left(\frac{dw}{w} - \frac{dr}{r}\right)}$$

$$= w\left[\frac{X_1C_{L1}C_{K1}\sigma_1}{(1 - a_1p_3)} + \frac{X_2C_{L2}C_{K2}\sigma_2}{(p_2 - a_2p_3)} + X_3C_{L3}C_{K3}\sigma_3\right] > 0$$

REFERENCES

R. N. Batra, "Protection and Real Wages Under Conditions of Variable Returns to Scale," *Oxford Econ. Pap.*, Nov. 1968, *20*, 353–60.

———, "Factor Accumulation and the Terms of Trade: A Three Country, Three Commodity, Three Factor Analysis," *Econometrica*, May 1970, *38*, 449–52.

——— and R. Singh, "Intermediate Products and the Two-Sector Growth Model," presented at the winter meetings, Econometric Society, Detroit, Dec. 1970.

J. N. Bhagwati and T. N. Srinivasan, "The Theory of Wage Differentials: Production Response and Factor Price Equalization," *J. Int. Econ.*, Feb. 1971, *1*, 19–35.

F. R. Casas, (1972a) "The Theory of Intermediate Products, Technical Change and Growth," *J. Int. Econ.*, May 1972, *2*, 189–200.

———, (1972b) "International Trade and Economic Growth in the Presence of Factor Market Distortions," *Amer. Econ.*, Spring 1972, *16*, 112–19.

E. F. Heckscher, "The Effect of Foreign Trade on the Distribution of Income," in H. S. Ellis and L. A. Metzler, eds., *Readings in the Theory of International Trade*, Philadelphia 1949.

R. W. Jones, "The Structure of Simple General Equilibrium Models," *J. Polit. Econ.*, Dec. 1965, *73*, 557–72.

———, "Variable Returns to Scale in General Equilibrium Theory," *Int. Econ. Rev.*, Oct. 1968, *9*, 261–72.

———, "Distortions in Factor Markets and the General Equilibrium Model of Production," *J. Polit. Econ.*, May/June 1971, *79*, 437–59.

M. C. Kemp, *The Pure Theory of International Trade and Investment*, New Jersey 1969.

——— and H. Herberg, "Factor Market Distortions, the Shape of the Locus of Competitive Outputs, and the Relation between Product Prices and Equilibrium Outputs," in J. N. Bhagwati et al., eds., *Trade, Balance of Payments, and Growth: Papers in International Economics in Honor of Charles P. Kindleberger*, Amsterdam 1971.

C. Khang, "A Dynamic Model of Trade between the Final and the Intermediate Products," *J. Econ. Theor.*, Dec. 1969, *1*, 416–37.

R. I. McKinnon, "Intermediate Products and Differential Tariffs: A Generalization of Lerner's Symmetry Theorem," *Quart. J. Econ.*, Nov. 1966, *80*, 584–615.

J. R. Melvin, "Production and Trade with Two Factors and Three Goods," *Amer. Econ. Rev.*, Dec. 1968, *58*, 1249–68.

———, "Intermediate Goods, the Production Possibility Curve, and Gains from Trade," *Quart. J. Econ.*, Feb. 1969, *83*, 141–51.

N. Minabe, "The Stolper-Samuelson Theorem, the Rybczynski Effect, and the Heckscher-Ohlin Theory of Trade Pattern and Factor Price Equalization: The Case of Many-Commodity, Many-Factor Country," *Can. J. Econ. Polit. Sci.*, Aug. 1967, *33*, 401–19.

B. Ohlin, *Interregional and International Trade*, Cambridge, Mass. 1933.

A. Ray, "Some Nontraded Inputs and Aspects of the Pure Theory of International Trade," presented to the Econometric Society meeting, New Orleans, Dec. 1971.

R. J. Ruffin, "Tariffs, Intermediate Goods, and Domestic Protection," *Amer. Econ. Rev.*, June 1969, *59*, 261–69.

T. M. Rybczynski, "Factor Endowment and Relative Commodity Prices," *Economica*, Nov. 1955, *22*, 336–41.

P. A. Samuelson, "Prices of Factors and Goods in General Equilibrium," *Rev. Econ. Stud.*, Oct. 1953, *20*, 1–20.

W. F. Stolper and **P. A. Samuelson,** "Protection and Real Wages," *Rev. Econ. Stud.*, Nov. 1944, *11*, 58–73.

J. Vanek, "Variable Factor Proportions and Interindustry Flows in the Theory of International Trade," *Quart. J. Econ.*, Feb. 1963, *77*, 129–42.

R. D. Warne, "Intermediate Goods in International Trade with Variable Proportions and Two Primary Inputs," *Quart. J. Econ.*, May 1971, *85*, 225–36.

P. L. Yates, *Forty Years of Foreign Trade*, London 1959.

The "F-Twist" and the Methodology of Paul Samuelson

By STANLEY WONG*

That economics is in an uneasy state is now widely recognized.[1] Undoubtedly, remedies to rectify this situation will involve methodological considerations about the role and status of theories in economics. Most economists will dread this prospect; most methodological discussions in economics are vague and confusing debates, which are of little apparent significance to the day-to-day work of the practicing economist. Regrettably, to this category of methodological debates must be added the celebrated methodological dispute between two of the most distinguished economists, Milton Friedman and Paul Samuelson.

What then is the excuse to write yet another paper on methodology? The alternative to bad methodological discussion is not *no* methodological discussion. As Samuelson put it quite succinctly:

> . . . a scholar in economics who is fundamentally confused concerning the relationship of definition, tautology, logical implication, empirical hypothesis, and factual refutation may spend a lifetime shadow-boxing with reality. In a sense, therefore, in order to earn his daily bread as a fruitful contributor to knowledge, the practitioner of an intermediately hard science like economics must come to terms with methodological problems. [1965b, p. ix]

It is the purpose of this paper to show that no substantive issues were at stake in the Friedman-Samuelson dispute. Many, perhaps most, economists consider the debate to be over, with Samuelson's position upheld.[2] The importance of sorting out the issues cannot be overemphasized. The methodological position we adopt profoundly influences our theoretical work and our assessment of it.

Section I provides a summary of Friedman's methodological position; Section II, Samuelson's critique (the F-Twist) of Friedman's position. Section III discusses Friedman's view as an example of instrumentalism. Section IV provides a critique of Samuelson's critique. Section V examines Samuelson's view as an example of descriptivism and argues that it is untenable. Finally, Section VI argues that the methodological choice is not one between the Friedman view and the Samuelson view *but* between both views and the view that theories are explanatory and informative.

I. The Methodology of Milton Friedman

In his provocative essay, "The Methodology of Positive Economics," Friedman gives the only detailed account of his methodology.

Following John Neville Keynes' distinction of normative and positive economics, Friedman argues that the role of

* King's College, Cambridge, England. The first version was presented in June 1971 to Joan Robinson's Seminar at the Faculty of Economics and Politics, University of Cambridge. This paper has benefitted enormously from the strong encouragement and penetrating criticisms of Joan Robinson, Lawrence Boland, G. C. Harcourt, Luigi Pasinetti, Alan Millar, Ian McFetridge, Gordon Nagel, Victoria Becker and Edwin Wong. The Canada Council has generously supported the research which made this paper possible.

[1] See articles by Frank Hahn and Wassily Leontief. Moreover, Joan Robinson (1964) has for years deplored the poverty of contemporary economic theory.

[2] Compare articles by D. V. T. Bear and Daniel Orr, Jack Melitz, and Tjalling Koopmans. My observation is shared by Karl Brunner.

the economist qua economist is to do positive economics, the study of "what is" rather than of "what should be." The task of positive economics ". . . is to provide a system of generalizations that can be used to make correct predictions about the consequences of any change in circumstances. Its performance is to be judged by the precision, scope, and conformity with experience of the predictions it yields" (p. 4).

How is a theory to be evaluated? For Friedman a ". . . theory is to be judged by its predictive power of the class of phenomena which it is intended to 'explain' . . . the only relevant test of the *validity* of a hypothesis is comparison of its predictions with experience" (pp. 8–9).

What about the descriptive realism of the assumptions or of the theory? Friedman claims that it is not possible to achieve since the enumeration of details need not stop in a description. He says that if complete descriptive realism were possible, it would not be desirable since it would render the theory utterly useless, devoid of any predictive power (see p. 32).

He asserts that the only relevant question to ask is ". . . whether they are sufficiently good approximations for the purpose at hand. And this question can be answered only by seeing whether a theory works, which means whether it yields sufficiently accurate predictions. The two supposedly independent tests [test of the assumptions and test of the theory by a test of its predictions] thus reduce to one test" (p. 15).

II. The F-Twist Methodology

Sensing the possible reemergence of the doctrine of *apriorism*, which he has scornfully attacked on many occasions (1952, p. 62; 1965b, p. 3), Samuelson assails Friedman's methodological view for praising the shortcomings of a theory as its virtues. According to Samuelson, Friedman's methodological position can best be summarized as the F-Twist: "A theory is vindicable if (some of) its consequences are empirically valid to a useful degree of approximation; the (empirical) unrealism of the theory 'itself' or of its 'assumptions,' is quite irrelevant to its validity and worth" (1963, p. 232).[3]

Samuelson's critique of Friedman's methodological position depends on what he (Samuelson) considers to be the proper relationship between a theory, its assumptions, and its consequences. Accordingly, a somewhat detailed account of Samuelson's critique follows.

Samuelson defines a theory (called B) ". . . as a set of axioms, postulates, or hypotheses that stipulate something about observable reality" (1963, p. 233), i.e., the set is either refutable or confirmable in principle by observation.[4] A theory has a set of consequences (called C) which are logically implied by the theory and a set of assumptions (called A) which logically implies the theory.

"If C is the complete set of consequences of B, it is identical with B . . . " (p. 234), and if A is the minimal set of assumptions that give rise to B, then A is identical with B. According to Samuelson, if C and A are given the above interpretation, they are logically equivalent to B and hence to each other. Therefore the degree of "realism," "factual correctness," "empirical validity," or "truth" of any one of A, B, C is shared by the other two.

Samuelson contends that it is a contradiction to maintain that any one of A, B, C can have a degree of realism different from the other two. He cites as an example of this point, Hendrik Houthakker's result (1950) that the ordinal theory of utility is

[3] Obviously, there is ambiguity in the usage of such words as theory, assumptions, implications, consequences, etc. This has hindered clarification of the issue.

[4] Samuelson does not seem to be aware of the problems associated with confirmation. For illuminating discussions of these problems, see John Watkins (1957), (1960) and Nelson Goodman.

logically equivalent to the Strong Axiom of revealed preference, that is, a rational consumer in conventional ordinal utility theory never violates the Strong Axiom.

In case only some part of A, B, C has empirical validity, Samuelson gives the following argument.

Consider a proper subset (called $C-$) of the consequence set C and let set A be a proper subset of the enlarged assumption set (called $A+$). Symbolically, he represents this relationship as[5]

(1) $A+\supset A\equiv B\equiv C\supset C-$

If C has "... complete (or satisfactory) empirical validity" (p. 234), then so does the theory, B, and the assumption, A. However, we cannot say anything about $A+$ "... unless its full content, which we may call $A+\equiv B+\equiv C+$ also have empirical validity. If that part of $C+$ which is not in C is unrealistic in the sense of being empirically false at the required level of approximation, then $A+$ is definitely the worse for it" (p. 234). According to Samuelson, then, it is absurd to maintain that if only some parts of C are valid, then B and A are important though invalid. The only thing to do he says is to eliminate that part of B and A corresponding to the invalid part of C and retain $A-\equiv B-\equiv C-$.

We can summarize Samuelson's critique as follows. 1) It is a contradiction to maintain that all consequences can be valid and the theory and the assumptions not valid. 2) It is absurd to maintain, in the case where only some of the consequences are valid, that the theory and the assumptions are important though invalid. The part of the theory set and the assumption set corresponding to the invalid part of the consequence set should be eliminated. For convenience let us call 1) the F-Twist Theorem and 2) the F-Twist Corollary.

[5] The $\supset$ is ambiguously used by Samuelson. It is not clear whether it signifies class inclusion or implication

III. Instrumentalism and the Methodology of Milton Friedman

Instrumentalism is the thesis that a theory in science is merely an instrument for prediction of observable reality.[6] Accordingly, a theory cannot be properly called true or false. It is tested only by the conformity of its predictions with observable reality. It is superfluous and irrelevant to test assumptions as they are merely tools or instruments which are judged by their ease or convenience in use. The adequacy of the assumptions cannot be decided independently of deciding whether the theory "works," i.e., whether the predictions are accurate enough. Furthermore, the testing of assumptions may mislead a theorist away from the testing of the theory by its predictions.

That Friedman is an instrumentalist is quite evident. The apparent ambiguities and inconsistencies in his essay can best be sorted out by considering his view as instrumentalist. All methodological prescriptions that Friedman makes are subsidiary to one overriding methodological maxim—that of successful prediction.

We can further elaborate on Friedman's methodological position by considering an alternative interpretation presented by Kurt Klappholz, an economist, and Joseph Agassi, a philosopher and historian of science, in their "Methodological Prescriptions in Economics."

In their review of Friedman's essay, Klappholz and Agassi contend that Friedman's position is essentially a critical one. However, they point out that by establishing certain methodological prescriptions for progress in economics he lapses occasionally into an uncritical one.

They illustrate their criticisms by examining three propositions in Friedman's essay:

[6] For a critical discussion of instrumentalism, see Karl Popper (1969).

(a) "A hypothesis can be tested *only* by the conformity of its implications or predictions with observable phenomena" (p. 40, our italics).
(b) "Great confidence is attached to .... a [hypothesis] if it has survived many opportunities for contradiction" (p. 9)....
(c) A new or rival theory "must have implications susceptible to empirical contradiction" before it can be regarded as interesting and important (p. 38). [p. 66]

Klappholz and Agassi challenge (a) as untrue because an untestable theory may be criticized, for example, by comparing it with another theory. Furthermore, they consider it to be dogmatic to lay such a heavy emphasis on testability. If Friedman's position is interpreted as instrumentalist, it can easily be seen that testability (in Friedman's sense) is both a necessary and a sufficient condition for the acceptability of a theory.

As for (b), they point out that absence of refutation does not mean one should have greater confidence in a theory. This proposition (b) can best be understood if we note Klappholz and Agassi's view of testing. For them, testing is a test of our understanding; for Friedman, it is a search for the limits of applicability. Once the limits are established, the instrumentalist will use the theory if it is successful in predicting events of a specified kind as determined by these limits. Confidence will be attached to a theory qua instrument if it is successful in making predictions on many occasions. A theory is not considered to be "contradicted" if it is refuted. Rather it is rejected ". . . if its predictions are contradicted ("*frequently*" or *more often* than predictions from alternative hypothesis)" (Friedman, p. 9, emphasis added). In other words, what Friedman means by noncontradiction is best understood in the context of his criterion of success of a theory—it works!

Klappholz and Agassi criticize (c) because it denies a yet unfalsifiable theory from careful consideration. Friedman's overriding concern for successful predictions rules out such careful consideration for fruitful alternative theories. Referring to the criticisms of economic theory by Thorstein Veblen and others, Friedman remarks that ". . . criticism of this type is largely beside the point unless supplemented by evidence that a hypothesis differing in one or another of these respects from the theory being criticized yields better predictions for as wide a range of phenomena" (p. 31).

Again we can understand Friedman's position. A theory must be the most successful predictor before it can be given serious consideration.

As a consequence of the instrumentalist position, allegedly false statements in the theory are not problematic if the theory can give sufficiently accurate predictions. Sufficiency is determined with respect to the predictive power of all other alternative theories. Similarly, statistical relationships can be postulated between variables without explaining the mechanism linking them provided such relationships are useful in generating predictions.

IV. A Critique of Samuelson's Critique

We shall now use the interpretation of Friedman's methodology given in Section III to construct a *possible* Friedman defense against Samuelson's critique.[7] In addition we shall, independent of instrumentalism, offer a critique of Samuelson's critique.

In Section II we summarized Samuelson's criticisms as

F-Twist Theorem: It is a contradiction

[7] It is not surprising that Friedman has not replied to his critics. Presumably, he regards arguing about methodology to be like arguing about realism of a theory or its assumptions.

to maintain that all consequences can be valid and the theory and the assumptions not valid.

F-Twist Corollary: It is absurd to maintain, in the case where only some of the consequences are valid, that the theory and the assumptions are important though invalid. The parts of the theory set and the assumption set corresponding to the invalid part of the consequence set should be eliminated.

First consider the Friedman-type defense against the critique. Friedman would certainly reject the Theorem since he does not accept Samuelson's view of the desired relationship between a theory and its assumption and consequences as one of logical equivalence. In fact, Friedman considers a logical equivalency as a vice (which we may call the "S-Twist"): ". . . economic theory must be more than a structure of tautologies if it is to be able to predict and not merely to describe the consequences of action" (p. 11).

Similarly Friedman would reject the Corollary since for his instrumentalism it is irrelevant that the theory set or its assumption set is invalid since what matters is whether the theory gives sufficiently accurate predictions for the purpose at hand.

Even if elimination of the part of the theory set or the assumption set corresponding to the invalid part of the consequence set were possible it may not always be desirable to do so for instrumentalism. The corresponding part of the theory set or the assumption set may be required for the generation of the valid predictions.

What then remains of Samuelson's critique? Perhaps, his claim that Friedman is praising a theory for its shortcomings, i.e., its unrealism? It is doubtful that Friedman intends this when he says, "Truly important and significant hypotheses will be found to have 'assumptions' that are wildly inaccurate descriptive representations of reality, and, in general, the more significant the theory, the more unrealistic the assumptions (in this sense)" (p. 20). For Friedman a theory is supposed to be an abstraction. It is the convenience of the simplicity produced by an abstraction which is its primary virtue, i.e., abstraction makes it possible for a theory to be used as an instrument for prediction. As noted in Section II, Friedman regards complete descriptive realism as a vice since that would render a theory useless for prediction (see Friedman, p. 15 and Ernest Nagel (1963)).

We must therefore conclude that Samuelson's misinterpretation of Friedman's methodology has led to the failure of his critique.[3]

It remains for us to consider whether Samuelson can still maintain both the Theorem and the Corollary. Unfortunately he cannot since both propositions are beset with numerous difficulties.

Let us first consider the Theorem. Samuelson says that if C is the *complete* or *full* set of consequences of the theory, B, then C is identical or logically equivalent to B and shares the same degree of realism with B.

First, it is not clear how completeness of C is shown. A theory is considered to be complete with respect to a set of statements if that set is logically deducible from the axiom set of the theory using the given

[3] Fritz Machlup charges Samuelson to be as much an "F-Twister" as Friedman. Machlup cites Samuelson's famous factor price-equalization theorem. Samuelson acknowledges that the list of assumptions of the theorem cannot be met in the real world but that " . . . a strong polar case like this . . . can often shed useful light on factual reality" (1964, p. 737). One wonders whether the F-Twist would be acceptable to Samuelson if it were reformulated as follows: If C is empirically valid (i.e., realistic at some degree of approximation), B and A are important provided they are realistic at some degree of approximation (the degree of approximation may not be the same as that of C, but it is determined by the successfulness of C). If C is successful, then the degree of approximation of B and A is acceptable.

rule(s) of inference.[9] Consequently, the complete set must be enumerated before showing whether the set is deducible from the axiom set. Secondly, allowing for the possibility that completeness can be determined, the argument still breaks down. Because a theory is shown to be complete, it does not mean that the derived propositions or theorems, Samuelson's C, are logically equivalent to the axiom set, except in the trivial sense that the axioms imply themselves. It is rather odd, to say the least, to prove the axiom set from the theorems since the axiom set is chosen such that the theorems can be proven. Of course, the possibility remains that by complete set of consequences Samuelson means the set that is logically equivalent to the theory set.

Thirdly, even if the set of axioms is logically equivalent to the set of consequences, his position is still untenable. For a theory to be explanatory it must satisfy a number of conditions. The set of statements which forms the explanation (called the *explanans*) must logically entail the set of statements which describe what is to be explained (called the *explanandum*). The explanans must not be known to be false and be independently testable. To be independently testable the explanans must have testable consequences in addition to the explanandum. Consequently, the explanans must be richer in explanatory power and empirical content than the explanandum. Moreover, to preclude *ad hoc* explanations, the explanans must also include at least one universal law.[10]

Clearly a statement of logical equivalence has no excess empirical content. The explanans (the axiom set) is just a restatement of the explanandum (the consequence set). There is no evidence for the explanans apart from that for the explanandum.[11] Thus in the theory of consumer behavior, used by Samuelson as an example of a successful economic theory, revealed preference theory is just a restatement of the ordinal utility theory (with certain regularity assumptions). Neither theory can be considered to be an explanation of the other.

Nevertheless, supporters of Samuelson's view would deny that theories are ever explanatory and that descriptions are all that are attainable. They consider the word "explanatory" to be an honorific title conferred on a better description. We shall take up this claim in Section V. The argument about his usage of the word "minimal" can be developed on similar lines.

Having cast sufficient doubt on the view that a theory is *ideally*[12] logically equivalent to its assumptions and its consequences, we have shown that the F-Twist Theorem is at minimum highly problematic. The possibility remains that the theory, its assumptions, and its consequences can have different degrees of realism or empirical validity. Of course it is wrong to say that the consequence set can be false and the axiom set, true. By *Modus Tollens*[13] the falsity of the consequence set,

[9] The meaning of completeness, chosen here, appears to be the most appropriate since the others (see Irving Copi, ch. 6) refer to more highly axiomatized theories than those found in economics. For a different view of completeness in economics, see Boland (1970b).

[10] For further discussion of the explanatory view of theories see Popper (1957), (1968), (1969).

[11] Most philosophers are in agreement on this point. See Nagel (1961) pp. 33–37; Popper (1968), ch. 3, especially Sec. 15, (1969), pp. 385–88; Carl Hempel, pp. 30–32; and Richard Braithwaite, ch. 3, especially p. 76.

[12] Samuelson seems to see logical equivalency as a theoretical goal. This point will be discussed in greater detail in Section V.

[13] *Modus Tollens* is a rule of inference in logic by which we can validly infer from the truth of a conditional statement and of the negation of the consequent clause of the conditional statement to the truth of the negation of the antecedent clause of the conditional statement. For example, from the truth of "If Socrates is a man, then he is mortal" and the truth of "It is not the case that he is mortal," we can validly infer by *Modus Tollens* the truth of "It is not the case that Socrates is a man." For further details see any standard textbook in logic, for example, see Copi, ch. 2.

i.e., of at least one consequence, implies the falsity of the axiom set.[14] That is, if at least one of the logical consequences has been shown to be false, then we know that at least one of the antecedents is false. Whether we can find the false antecedent(s) is a difficult question to answer.[15] However, to make the above criticism of an invalid argument, we need not hold the view that a theory is logically equivalent to its consequences.

Let us now turn to the Corollary, which is closely linked to the Theorem. Samuelson argues that it is absurd to claim that a theory and its assumptions can be important though invalid if some of the consequences are invalid. He suggests that the invalid parts of the theory and of the assumptions corresponding to the invalid consequences should be eliminated.

The first part of the Corollary follows from the Theorem. On its own what it means is not clear. As to the call for the elimination of the parts of the theory and assumptions corresponding to the invalid consequences, Samuelson fails to recognize that it is not always possible to identify the corresponding parts. Furthermore, if the parts could be so identified it may not be possible to eliminate them if they are required for the explanation and prediction of the valid or true consequences. (See also fn. 14.)

We have now shown that there is not much substance to Samuelson's critique from an instrumentalist point of view. Moreover, we have shown, independently of instrumentalism, that Samuelson's critique is beset with numerous difficulties.[16]

V. Descriptivism and the Methodology of Paul Samuelson[17]

The following discussion can only be a preliminary examination of the methodology of Samuelson. Unfortunately, no account of the methodology of such an influential economist has yet appeared in economic literature. In this section attention will be given to his writings on methodological issues. A more thorough examination would require, in addition, a consideration of his methodological views that are either implicit or explicit in his writings on theoretical and policy issues.

In Section IV, it was shown that Samuelson's presentation of the relationship between a theory and its assumptions and consequences is wrong since for a theory to be explanatory and informative, it must go beyond redescription of what is to be explained.[18]

As noted in Section IV, Samuelson rejects the view that theories are explanatory. "Every description that is superseded by a 'deeper explanation' turns out upon careful examination to have been replaced by still another description, albeit possibly

[14] On falsification see Popper (1968), ch. 1, especially Sec. 6 and ch. 5.

[15] See W. V. Quine, " . . . our statements about the external world face the tribunal of sense experience not individually but only as a corporate body," (p. 41, see also pp. 42–46). See also Lakatos, Appendix.

[16] For those who do not accept the dismissal of Samuelson's critique, there is the option to criticize his meta-assumption(s) rather than his meta-conclusions, the F-Twist Theorem and the F-Twist Corollary. Now, this option should be welcomed by Samuelson as it is an important aspect of his critique of Friedman's rejection of criticism of assumptions. This option will be taken up in Section V. If we are successful in criticizing his meta-assumption(s), Samuelson must cease using the Theorem and the Corollary. Should the defenders or advocates of Samuelson's critique continue to assert them, that is, the same meta-conclusions regardless of Samuelson's meta-assumption(s), they will be caught in a contradiction. Samuelson's defenders will then be practicing Friedman's instrumentalist methodology at the meta-level.

[17] The views of Samuelson and Friedman are not mutually exclusive. In fact, Boland (1970a), (1971) maintains that they both share the same methodology, that of conventionalism. The debate is then seen as a family dispute over which conventionalist criteria has priority, generality or simplicity.

[18] By go beyond redescription, it is not suggested that a theory is therefore necessarily a false description. What is advocated is that a theory to be informative must not be a restatement of what is to be explained. Samuelson (1965a) does not seem to understand this point which was made by Gerald Massey.

a more useful description that covers, and illuminates a wider area" (1964, p. 737). An explanation is just an honorific title for a better description, that is, "*. . . a better kind of description and not something that goes ultimately beyond description*" (1965a, p. 1165; see also 1965c, p. 103).

Why does Samuelson regard a theory to be only a description and not an explanation? Why does he *want* a theory to be logically equivalent to its assumptions and to its consequences?

The answers to both questions are rooted in Samuelson's particular methodological view, a version of *descriptivism* (see Nagel (1961), especially pp. 118–29).[19] The answers given will be explanations which are not ultimate but remain always conjectural. Nevertheless, they offer reasons for Samuelson's choice of methodological position.

According to Samuelson's descriptivist methodology, 1) a theory is just a description of observable experience, a convenient and mnemonic representation of empirical reality (1952, p. 61); (1963, p. 236); (1965a, p. 1171). As a corollary, 2) knowledge consists essentially of observational reports; a theory expressible in observational language is superior to those that are not. The reasons for rejecting the view that theories are explanatory are: 3) Explanations are ultimate (1965c, pp. 102–03). 4) Apriorism must be avoided, hence theories should be expressed in observational language (1963, p. 235); (1964, p. 738). 5) Explanations turn out to be just better descriptions (1964, p. 737); (1965a, p. 1165). 6) All well-known theories in science are expressible in terms of observational statements, i.e., basic statements (1965a, p. 1167).

With the above interpretation of Samuelson's methodology, we can now see why Samuelson has logical equivalency as a theoretical goal. Since for him the object of a theory is to represent observable experiences, he considers as a successful theory one which is logically equivalent to statements of facts, the observational report. Moreover, he says that a theory is approximate, and worse off because of it, if it is not equivalent to statements of the observable reality.

Given the goal of logical equivalency, we can now consider Samuelson's view of the development of theory in science, including economics.

We can represent Samuelson's schema below. It is adapted from his 1963 paper. Each entry is a proper subset (or consequent, since the notation is used ambiguously by Samuelson) of the entry above it. In the northerly direction, there is increasing generality or universality. At each level, an attempt is made to establish a logical equivalency between a theory and its assumptions and its consequences. Once this relationship is achieved, the theoretical task is completed. A search is then made to expand the framework by another logical equivalency between a theory and its assumptions and consequences at the next level up. Samuelson has viewed as an example of successful theoretical work, the theory of consumer's behavior, because the conventional ordinal utility theory (B) was shown by Houthakker to be equivalent to the revealed preference approach (C), i.e., $B \equiv C$.

$$
\begin{array}{llll}
 & \textit{Assumptions} & \textit{Theory} & \textit{Consequences} \\
 & A+\ldots+ & \equiv B+\ldots+ & \equiv C+\ldots+ \\
\uparrow & \vdots & \vdots & \vdots \\
 & \supset & \supset & \supset \\
 & A+ & \equiv B+ & \equiv C+ \\
 & \supset & \supset & \supset \\
 & A & \equiv B & \equiv C \\
\text{Generality} & \supset & \supset & \supset \\
 & A- & \equiv B- & \equiv C- \\
 & \vdots & \vdots & \vdots
\end{array}
$$

[19] The term descriptivist is used rather than the usual descriptive because the view attacked is not that theories are descriptive but that they are *only* descriptive.

Now a refutation of Samuelson's descriptivist methodology will be given. With such a refutation, we can reject the goal of logical equivalency since it is just a specification of his descriptivist methodology.

Explanations are not ultimate. We can give explanations of explanations. Samuelson mistakenly identifies the explanatory view with an essentialist view. *Essentialism* aims at ultimate explanation which is neither capable nor in need of further explanation. It maintains that a theory goes behind the appearance of phenomena to reveal their underlying essences (see Popper (1969)). Clearly, essentialism is not a necessary part of the explanatory view.

Samuelson has rightfully condemned apriorism, the view that all phenomena can be explained as consequences of self-evident first principles, the truth of which is independent of all possible experience. But the alternative to such apriorism is not to ground theories in observational statements. For such a program is impossible to achieve because of insurmountable difficulties, both logical and epistemological.

First, the logical one, a theory is not logically equivalent to a set of observational statements because of the logical form of a theory. An observational statement, that is, a basic statement, is of the form, "There exists an x with property P at place-time r, t." A theory on the other hand includes at least one unrestricted universal statement of the form, "For all x, if x is P, then x is Q," where no spatio-temporal location is specified. A theory, therefore, cannot be logically equivalent to a set of observational statements because an unrestricted universal statement is *not* equivalent to a finite conjunction of observational statements, i.e., basic statements.

Samuelson's contention that well-known theories in science are expressible in terms of basic sentences (statements) is wrong. This contention was made in response to Massey's remarks that ". . . empirical science comes to maturity only after it effects a clean break with basic sentences, only after it boldly postulates theoretical statements that ultimately are anchored, though not submerged, in experience by means of semantic ties to basic sentences" (p. 1163).[20]

Charging Massey with using "imperialistic" vocabulary to refer to indirectly observable sentences, Samuelson claims that to accept his views would be tantamount to considering well-known theories in science to be inferior since, these too, are expressible in terms of Massey's basic sentences. Claiming the support of several physicists, Samuelson maintains that the following theories are all expressible in terms of basic sentences: Galileo's theory of falling bodies, Newtonian theory of gravitation as applied to the n-body problem, Einstein's special theory of relativity, and classical thermodynamics of Carnot, Clausius, Kelvin and pre-1900 Gibbs. However, the examples he cites cannot be so expressed. All these theories claim to be universal. For example, Galileo's theory does not apply only to a finite number of bodies but to any body. Therefore, they are not equivalent to a finite conjunction of basic sentences.

Secondly, there does not exist an independent observational language in which to ground theories and theoretical concepts. While one may accept the criticism that a theory cannot be expressed as a

[20] Much of the confusion evident in the exchange between Samuelson and Massey can be attributed to Massey's rather uncritical use of the term basic. Massey considers those sentences which are basic to be properly called true or false. Theoretical sentences can not be called true or false; their acceptance is only indirectly linked with observable experience. To draw such a distinction between theoretical and basic sentences, Massey has adopted the view that basic sentences are the fundamental and thus indisputable building blocks of scientific theories. This view overlooks the theoretical content of basic sentences and thus that theoretical sentences can be true or false.

finite conjunction of basic sentences, one still may wish to maintain that a theory is observational in so far as it contains observational terms. For example, "For all x, if x is P, then x is Q" may be considered to be observational because P and/or Q are observational terms. Unfortunately, this view cannot be maintained without encountering further difficulties. All observational terms are theory-laden. They implicitly presume theories which by their very nature go beyond just redescription of the facts. These theories prescribe lawlike behavior to certain entities. For example, such observational terms as glass or water presume lawlike behavior. We can challenge whether a container is a glass by testing the container for properties of a glass.[21]

To accept the view that all observational terms are theory-laden would be incompatible with the descriptivist position. For the view that knowledge consists essentially of observational reports is incompatible with the view that all observational terms are theory-laden. If the informative part of a statement is the observational, then we do not know what a statement is asserting unless we can separate the observational from the theoretical. But then we cannot do this.[22]

Explanations are not just better descriptions. This assertion can be supported through an examination of Samuelson's claims that Newton's theory is just a better description than Kepler's theory (1964, p. 737) and that Kepler's three laws together are necessary and sufficient conditions for the truth of Newton's theory (1965a, p. 1169). Both of these claims are false.

That Kepler's results are good approximations to Newton's is not denied. But only after Newton's theory was established were Kepler's results corrected. Further, Newton sought to explain (and consequently, corrected) Kepler's results by deducing them from more fundamental laws, the Laws of Motion and Universal Gravitation. Therefore, to say that explanations are just better descriptions is unfounded historically.

To deduce Kepler's third law from Newtonian theory, we have to assume that all planets are of the same mass. But that assumption is false. Therefore, Kepler's laws are not necessary and sufficient conditions for the truth of Newton's theory (see Max Born, pp. 129–33; Popper (1957, p. 29 and passim); Pierre Duhem, Part II, ch. 6, Section 4; and Herbert Goldstein, p. 80). Thus the refutation of Samuelson's descriptivism is completed.

With the refutation of the descriptivist view and thus, of the goal of logical equivalency we can now understand the significance of Samuelson's claim that a success was achieved when Houthakker showed the equivalence of the conventional utility approach and the revealed preference approach. Houthakker in his well-known survey of consumer theory adopts the same view as Samuelson:

> The principal qualification to Friedman's methodological position that needs to be made is that in a well-developed theory the distinction between assumptions and conclusions disappears. If a theory reaches logical (as opposed to empirical) perfection, the assumptions under which any conclusion is derived are just strong enough for that purpose; in other words the assumptions are necessary and sufficient for the conclusions, and each may be judged by the other. The theory then becomes a list of equivalent relations between two sets of concepts, neither one of which is more fundamental than the other.
>
> [1961, p. 705]

[21] See Popper (1968, especially Appendix *x) for a more detailed discussion of this point. For a somewhat different view of theory-ladenness see Kuhn and for a critical assessment of Kuhn's view see articles in Lakatos and Musgrave.

[22] For a devastating criticism of sensationalism of which descriptivism is a part, see Agassi.

As noted earlier, if we show that a theory is logically equivalent to its assumptions and to its consequences, then we must reject the theory as an *explanation* of the consequences and the assumptions as an *explanation* of the theory. A set of statements used to explain another set of statements cannot itself be explained by the other set of statements. One may ask what purpose is served by choosing to group statements under the categories: assumptions, theory, and consequences. For Samuelson's descriptivism, the choice has no epistemological basis, only a conventional (arbitrary) one.

Samuelson cites as an example of a well-developed theory, the theory of consumer behavior: "The doctrines of revealed preference provide the most literal example of a theory that has been stripped down to its bare implications for empirical realism: Occam's Razor has cut away every zipper, collar, shift, and fig leaf" (1964, p. 738).

Referring to the work of Houthakker on revealed preference, Samuelson asserts ". . . that the conventional theory has *no* wider implications than these prosaic factual implications [i.e., one never observes the contradiction of the Strong Axiom]. Once the two formulations had been rigorously proved to imply each other mutually, that issue was settled" (1964, p. 738).

These comments on the development of the theory of consumer's behavior are rather curious. The theory of revealed preference was originally offered as a replacement for the ordinal utility theory because Samuelson wished to rid the theory of consumer behavior of all dependence on empirically untestable hedonistic psychology and utilitarian philosophy (1938), (1965b). The Strong and Weak Axioms were considered to be the empirically testable versions of the theory of consumer behavior. With the work of Houthakker (1950) and Hirofumi Uzawa we now know that the two versions of the theory are logically equivalent. Hence, the revealed preference version is no more testable than the ordinal utility version. In fact, no desired goals appear to be achieved by this new approach. If so, why does Samuelson still advocate the pursuit of logical equivalencies?

That Samuelson seeks logical equivalencies is evident by considering his view of the role of deduction: ". . . it is clear that deduction has the modest linguistic role of translating certain empirical hypotheses into their 'logical equivalents' " (1952, p. 57; see also 1965b, p. 12). In terms of the analysis in Section IV, the logical equivalent of an empirical hypothesis (theory) is presumably the complete consequence set. Samuelson's view of the *role* of deduction seems rather unusual. This view of the role of deduction is only compatible with the goal of logical equivalency.

Nevertheless, Samuelson insists that we *always* derive logical equivalents in deduction:

> Actually, if proposition A correctly implies proposition B, and B correctly implies proposition C, and so forth all the way to Z, then it is necessarily true that A implies Z in every sense that it implies B. There can be no leakage of truth at any stage of a valid deductive syllogism. All such syllogisms are mere translations of the type, "A rose is a rose is a rose." [1952, p. 58]

A implies Z does not mean that Z is logically equivalent to A. For A to be logically equivalent to Z, we must show that A implies Z and Z implies A.

While we can, if we so wish, regard the *role* of deduction as one of deducing logical equivalents from empirical hypotheses, it is wrong to assert that *all* logical consequents are logical equivalents.

VI. What Is Not the Methodological Choice?

While the paper has been primarily directed to the refutation of Samuelson's

espoused methodology, it should now be clear that the refutation should not be interpreted as a vindication of Friedman's position.

Throughout the debate in economic literature the choice has been between either Friedman's or Samuelson's position. Unrealism is regarded as a Samuelson vice (the F-Twist) but in some sense a Friedman virtue (in view of instrumentalism); pure description, the end of which is a logical equivalency, a Friedman vice (which in Section IV we have called the S-Twist) but a Samuelson virtue (in view of descriptivism). Descriptivism was shown to be untenable in view of logical and epistemological difficulties but instrumentalism cannot be defeated in a similar fashion.

In fact no refutation appears possible. Although Samuelson's descriptivism does not allow a valid critique of Friedman's instrumentalism, one can appreciate the spirit of Samuelson's F-Twist caricature, namely, that Friedman's methodology is fundamentally anti-intellectual. Instrumentalism provides a self-justification for its methods and its methodological choices. This self-justification is quite apparent in view of the current controversy about the *new monetarism* over such issues as the relation of the money supply to the money *GNP*. Harry Johnson attributes in part the rise of the new monetarism to Friedman's methodology: ". . . the methodology of positive economics was an ideal methodology for justifying work that produced apparently surprising results without feeling obliged to explain just why they occurred and in so doing mystifying and exciting the interest of noncommitted economists and wavering Keynesians" (p. 13).

However, he correctly notes the limitation of instrumentalism and the difficulty involved in defeating it: "The demand for clarification of the mechanism by which results can be explained is contrary to the methodology of positive economics with its reliance on the 'as if' approach" (p. 13).

Why does it matter if we choose one methodology or another? Why should we even be *self-conscious* of our methodological choices?

Although perhaps not immediately obvious, many economic controversies are at their roots, methodological ones.[23] In addition to the dispute over the new monetarism, explicit methodological considerations may throw light on a controversy which has enthralled the leading economic theoreticians for over twenty years, the controversy over the theory of capital and growth.[24]

We can briefly consider two recent results in the raging controversy. In challenging Robert Solow's claim of the importance of Irving Fisher's concept of the "rate of return" to the determination of the "rate of profit," Luigi Pasinetti points out:

> After *calling* "social rate of return" the rate of profit at which two economic systems are equally profitable one should not be surprised to find that the social rate of return so defined is indeed, in all circumstances, equal to the rate of profit at which the two economic systems are equally profitable. This has become a tautological assertion, which no doubt is always true but which cannot prove anything on any theory of the rate of profit. [1969, p. 526]

And furthermore, he points out that one may *choose* to use the words the "rate of return" as a synonym for the "rate of profit" but that one cannot ". . . obviously pretend to use the notion of rate of return to *explain* the rate of profit" (p. 529, see also (1970)).

Solow retorts that it is silly to call it tautological: "*Any* deductive proposition

[23] With Houthakker's result, one would have expected a controversy to develop in the theory of consumer behavior.

[24] For a detailed exposition of the main issues, see G. C. Harcourt.

is tautological; one can only say that some are less obvious or less interesting than others" (p. 425, emphasis added). It would appear that Solow shares the same view as Samuelson on the role of deduction and the role of theories.

Finally, consider the recent results on the usefulness of the Cobb-Douglas production function to *explain* the relative constancy of shares going to labor and capital. Commenting on the measure of capital controversy, Joan Robinson reports on the apparent success of the Cobb-Douglas function:

> The suggestion is clear, however, that labor's share is not roughly constant because the diverse technical relationships of modern economies are truly representable by an aggregate Cobb-Douglas but rather that such relationships appear to be representable by an aggregate Cobb-Douglas *because* labor's share happens to be roughly constant. [F. Fisher, *Rev. Econ. Statist.*, Nov. 1971, as quoted in Robinson (1971, p. 602).]

Of course, the way is open for the rebuttal that the Cobb-Douglas is not really to *explain* but to *describe* the empirical relationships found in economies.

The above examples are suggestive of the possibilities open to explicit methodological considerations in sorting out the issues at stake in economic controversies.

What then is the methodological choice? It surely is not instrumentalism or descriptivism. Instrumentalism in its single-minded pursuit of predictions *goes* "beyond the facts" by considering the truth or falsity of statements to be irrelevant; descriptivism in its pursuit of pure descriptions designs a theory *not to go* beyond the facts and thus ends up with a theory being just a restatement of the "facts."

As an alternative to the two views, we have advocated the view that a theory is explanatory and informative, namely, that in order to have any explanatory content it must go beyond pure description.

Hence, the descriptivist goal of logical equivalency denies a theory any explanatory content while the instrumentalist's particular obsession with predictions disregards the explanatory content of a theory.

The choice, then, is *not* between instrumentalism and descriptivism but between them *both* and the view that a theory is explanatory and informative, one which provides an answer, albeit a tentative one, to the question, "Why?"

REFERENCES

J. Agassi, "Sensationalism," *Mind*, Jan. 1966, *75*, 1–24.

D. Bear and D. Orr, "Logic and Expediency in Economic Theorizing," *J. Polit. Econ.*, Apr. 1967, *75*, 188–96.

L. Boland, "Economic Understanding and Understanding Economics," *S. Afr. J. Econ.*, June 1969, *37*, 144–60.

———, (1970a) "Conventionalism and Economic Theory," *Philosophy of Science*, June 1970, *37*, 239–48.

———, (1970b) "Axiomatic Analysis and Economic Understanding," *Australian Econ. Pap.*, June 1970, *9*, 62–75.

———, "Methodology as an Exercise in Economic Analysis," *Philosophy Sci.*, Mar. 1971, *38*, 105–17.

M. Born, *Natural Philosophy of Cause and Chance*, Oxford 1949.

R. Braithwaite, *Scientific Explanation*, Cambridge 1953.

K. Brunner, " 'Assumptions' and the Cognitive Quality of Theories," *Synthese*, Dec. 1969, *20*, 501–25.

I. Copi, *Symbolic Logic*, 3d edition, New York 1967.

P. Duhem, *The Aim and Structure of Physical Theory*, New York 1962.

M. Friedman, "The Methodology of Positive Economics," *Essays in Positive Economics*, Chicago 1953, 3–43.

G. Garb, "Professor Samuelson on Theory and Realism: Comment," *Amer. Econ. Rev.*, Dec. 1965, *55*, 1151–53.

H. Goldstein, *Classical Mechanics*, Reading 1950.

N. Goodman, *Fact, Fiction, and Forecast*, Indianapolis 1965.

F. **Hahn**, "Some Adjustment Problems," *Econometrica*, Jan. 1970, *38*, 1–17.

G. **Harcourt**, *Some Cambridge Controversies in the Theory of Capital*, Cambridge 1972.

C. **Hempel**, *Philosophy of Natural Science*, Englewood Cliffs 1966.

H. **Houthakker**, "Revealed Preference and the Utility Function," *Economica*, May 1950, *17*, 159–74.

———, "The Present State of Consumption Theory," *Econometrica*, Oct. 1961, *29*, 704–40.

H. **Johnson**, "The Keynesian Revolution and the Monetarist Counter-Revolution," *Amer. Econ. Rev. Proc.*, May 1971, *61*, 1–14.

K. **Klappholz** and J. **Agassi**, "Methodological Prescriptions in Economics," *Economica*, Feb. 1959, *26*, 60–74.

T. **Koopmans**, *Three Essays on the State of Economic Science*, New York 1957.

T. **Kuhn**, *The Structure of Scientific Revolutions*, Chicago 1970.

I. **Lakatos**, "Falsification and the Methodology of Scientific Research Programmes," in I. Lakatos and A. Musgrave, eds., *Criticism and the Growth of Knowledge*, Cambridge 1970, 91–195.

——— and A. **Musgrave**, *Criticism and the Growth of Knowledge*, Cambridge 1970.

W. **Leontief**, "Theoretical Assumptions and Nonobserved Facts," *Amer. Econ. Rev.*, Mar. 1971, *61*, 1–7.

A. **Lerner**, "Professor Samuelson on Theory and Realism: Comment," *Amer. Econ. Rev.*, Dec. 1965, *55*, 1153–55.

F. **Machlup**, "Professor Samuelson on Theory and Realism," *Amer. Econ. Rev.*, Sept. 1964, *54*, 733–36.

G. **Massey**, "Professor Samuelson on Theory and Realism: Comment," *Amer. Econ. Rev.*, Dec. 1965, *55*, 1155–64.

J. **Melitz**, "Friedman and Machlup on the Significance of Testing Economic Assumptions," *J. Polit. Econ.*, Feb. 1965, *73*, 37–60.

E. **Nagel**, *The Structure of Science*, London 1961.

———, "Assumptions in Economic Theory," *Amer. Econ. Rev. Proc.*, May 1963, *53*, 211–19.

L. **Pasinetti**, "Switches of Technique and the 'Rate of Return' in Capital Theory," *Econ. J.*, Sept. 1969, *79*, 508–531.

———, "Again on Capital Theory and Solow's 'Rate of Return'," *Econ. J.*, June 1970, *80*, 428–31.

K. **Popper**, "The Aim of Science," *Ratio*, Dec. 1957, *1*, 24–35.

———, *The Logic of Scientific Discovery*, London 1958.

———, "Three Views Concerning Human Knowledge," *Conjectures and Refutations*, London 1969, 97–119.

W. **Quine**, "Two Dogmas of Empiricism," *From a Logical Point of View*, New York 1961, 20–46.

J. **Robinson**, *Economic Philosophy*, Harmondsworth, Middlesex 1964.

———, "The Measure of Capital: The End of the Controversy," *Econ. J.*, Sept. 1971, *81*, 597–602.

P. **Samuelson**, "A Note on the Pure Theory of Consumer's Behavior," *Economica*, Feb. 1938, *5*, 61–71.

———, "Economic Theory and Mathematics—An Appraisal," *Amer. Econ. Rev. Proc.*, May 1952, *42*, 56–66.

———, "Problems of Methodology—Discussion," *Amer. Econ. Rev. Proc.*, May 1963, *53*, 231–36.

———, "Theory and Realism: A Reply," *Amer. Econ. Rev.*, Sept. 1964, *54*, 736–39.

———, (1965a) "Professor Samuelson on Theory and Realism: Reply," *Amer. Econ. Rev.*, Dec. 1965, *55*, 1164–72.

———, (1965b) *Foundations of Economic Analysis*, New York 1965.

———, (1965c) "Some Notions on Causality and Teleology in Economics," in D. Lerner, ed., *Cause and Effect*, New York 1965, 99–143.

R. **Solow**, "On the Rate of Return: Reply to Pasinetti," *Econ. J.*, June 1970, *80*, 423–28.

H. **Uzawa**, "Preference and Rational Choice in the Theory of Consumption," in K. Arrow et al., eds., *Proceedings of a Symposium on Mathematical Methods in the Social Sciences*, Stanford 1960, 129–48.

J. **Watkins**, "Between Analytic and Empirical," *Philosophy*, Apr. 1957, *32*, 112–31.

———, "When are Statements Empirical?" *British J. Philosophy Sci.*, Feb. 1960, *10*, 287–308.

Some International Evidence on Output-Inflation Tradeoffs

By Robert E. Lucas, Jr.*

This paper reports the results of an empirical study of real output-inflation tradeoffs, based on annual time-series from eighteen countries over the years 1951–67. These data are examined from the point of view of the hypothesis that average real output levels are invariant under changes in the time pattern of the rate of inflation, or that there exists a "natural rate" of real output. That is, we are concerned with the questions (i) does the natural rate theory lead to expressions of the output-inflation relationship which perform satisfactorily in an econometric sense for all, or most, of the countries in the sample, (ii) what testable restrictions does the theory impose on this relationship, and (iii) are these restrictions consistent with recent experience?

Since the term "natural rate theory" refers to varied aggregation of models and verbal developments,[1] it may be helpful to sketch the key elements of the particular version used in this paper. The first essential presumption is that *nominal* output is determined on the aggregate demand side of the economy, with the division into real output and the price level largely dependent on the behavior of *suppliers* of labor and goods. The second is that the partial "rigidities" which dominate short-run supply behavior result from suppliers' lack of information on some of the prices relevant to their decisions. The third presumption is that inferences on these relevant, unobserved prices are made optimally (or "rationally") in light of the stochastic character of the economy.

As I have argued elsewhere (1972), theories developed along these lines will *not* place testable restrictions on the coefficients of estimated Phillips curves or other single equation expressions of the tradeoff. They will not, for example, imply that money wage changes are linked to price level changes with a unit coefficient, or that "long-run" (in the usual distributed lag sense) Phillips curves must be vertical. They *will* (as we shall see below) link supply parameters to parameters governing the stochastic nature of demand shifts. The fact that the implications of the natural rate theory come in this form suggests an attempt to test it using a sample, such as the one employed in this study, in which a wide variety of aggregate demand behavior is exhibited.

In the following section, a simple aggregative model will be constructed using the elements sketched above. Results based on this model are reported in Section II, followed by a discussion and conclusions.

I. An Economic Model

The general structure of the model developed in this section may be described very simply. First, the aggregate price-quantity observations are viewed as intersection points of an aggregate demand and an aggregate supply schedule. The former is drawn up under the assumption of a cleared money market and represents the output-price level relationship implicit in

* Graduate School of Industrial Administration, Carnegie-Mellon University.

[1] The most useful, general statements are those of Milton Friedman (1968) and Edmund Phelps. Specific illustrative examples are provided by Donald Gordon and Allan Hynes and Lucas (April 1972).

the standard IS-LM diagram. It is viewed as being shifted by the usual set of demand-shift variables: monetary and fiscal policies and variation in export demands. The supply schedule is drawn under the assumption of a cleared labor market; its slope therefore reflects labor and product market "rigidities."

The structure of this model, which is essentially that suggested in Lucas and Leonard Rapping (1969), will be greatly simplified by an additional special assumption: that the aggregate demand curve is unit elastic.[2] In this case, the level of nominal output can be treated as an "exogenous" variable with respect to the goods market, and the entire burden of accounting for the breakdown of nominal income into real output and price is placed on the aggregate supply side. In the next subsection, *A*, a supply model designed to serve this purpose is developed. In subsection *B*, solutions to the full (demand and supply) model are obtained.

A. Aggregate Supply

All formulations of the natural rate theory postulate rational agents, whose decisions depend on *relative* prices only, placed in an economic setting in which they cannot distinguish relative from general price movements. Obviously, there is no limit to the number of models one can construct where agents are placed in this situation of imperfect information; the trick is to find tractable schemes with this feature. One such model is developed below.

We imagine suppliers as located in a large number of scattered, competitive markets. Demand for goods in each period is distributed unevenly over markets, leading to relative as well as general price movements. As a consequence, the situation as perceived by individual suppliers will be quite different from the aggregate situation as seen by an outside observer. Accordingly, we shall attempt to keep these two points of view separate, turning first to the situation faced by individual suppliers.

Quantity supplied in each market will be viewed as the product of a normal (or secular) component common to all markets and a cyclical component which varies from market to market. Letting z index markets, and using y_{nt} and y_{ct} to denote the *logs* of these components, supply in market z is:

$$y_t(z) = y_{nt} + y_{ct}(z) \tag{1}$$

The secular component, reflecting capital accumulation and population change, follows the trend line:

$$y_{nt} = \alpha + \beta t \tag{2}$$

The cyclical component varies with perceived, *relative* prices and with its own lagged value:

$$y_{ct}(z) = \gamma[P_t(z) - E(P_t \mid I_t(z))] + \lambda y_{c,t-1}(z) \tag{3}$$

where $P_t(z)$ is the actual price in z at t and $E(P_t \mid I_t(z))$ is the mean current, general price level, conditioned on information available in z at t, $I_t(z)$.[3] Since y_{ct} is a deviation from trend, $|\lambda| < 1$.

[2] An explicit derivation of the price-output relationship from the IS-LM framework is given by Frederic Raines. Of course, this framework does not imply an elasticity of unity, though it is consistent with it. Since the unit elasticity hypothesis is primarily a matter of convenience in the present study, I shall comment below on the probable consequences of relaxing it.

[3] A supply function for labor which varies with the ratio of actual to expected prices is developed and verified empirically by Lucas and Rapping (1969). The effect of lagged on actual employment is also shown. In our 1972 paper, in response to Albert Rees's criticism, we found that this persistence in employment cannot be fully explained by price expectations behavior. Both these effects—an expectations and a persistence effect—will be transmitted by firms to the goods market. In addition, they are probably augmented by speculative behavior on the part of firms (as analyzed for example, by Paul Taubman and Maurice Wilkinson).

For a general equilibrium model in which suppliers behave essentially as given by (3), see my 1972 papers.

The information available to suppliers in z at t comes from two sources. First, traders enter period t with knowledge of the past course of demand shifts, of normal supply y_{nt}, and of past deviations $y_{c,t-1}$, $y_{c,t-2}$, ... While this information does not permit exact inference of the *log* of the current general price level, P_t, it does determine a "prior" distribution on P_t, common to traders in all markets. We assume that this distribution is known to be normal, with mean $\bar{P}_t$ (depending in a known way on the above history) and a constant variance σ^2.

Second, we suppose that the actual price deviates from the (geometric) economy-wide average by an amount which is distributed independently of P_t. Specifically, let the percentage deviation of the price in z from the average P_t be denoted by z (so that markets are indexed by their price deviations from average) where z is normally distributed, independent of P_t, with mean zero and variance τ^2. Then the observed price in z, $P_t(z)$ (in *logs*) is the sum of independent, normal variates

$$P_t(z) = P_t + z \tag{4}$$

The information $I_t(z)$ relevant for estimation of the unobserved (by suppliers in z at t) P_t, consists then of the observed price $P_t(z)$ and the history summarized in $\bar{P}_t$.

To utilize this information, suppliers use (4) to calculate the distribution of P_t, conditional on $P_t(z)$ and $\bar{P}_t$. This distribution is (by straightforward calculation) normal with mean:

$$\begin{aligned} E(P_t \mid I_t(z)) &= E(P_t \mid P_t(z), \bar{P}_t) \\ &= (1-\theta)P_t(z) + \theta\bar{P}_t \end{aligned} \tag{5}$$

where $\theta=\tau^2/(\sigma^2+\tau^2)$, and variance $\theta\sigma^2$. Combining (1), (3), and (5) yields the supply function for market z:

$$\begin{aligned} y_t(z) = y_{nt} &+ \theta\gamma[P_t(z) - \bar{P}_t] \\ &+ \lambda y_{c,t-1}(z) \end{aligned} \tag{6}$$

Averaging over markets (integrating with respect to the distribution of z) gives the aggregate supply function:

$$\begin{aligned} y_t = y_{nt} &+ \theta\gamma(P_t - \bar{P}_t) \\ &+ \lambda[y_{t-1} - y_{n,t-1}] \end{aligned} \tag{7}$$

The *slope* of the aggregate supply function (7) thus varies with the fraction θ of *total* individual price variance, $\sigma^2+\tau^2$, which is due to *relative* price variation. In cases where τ^2 is relatively small, so that individual price changes are virtually certain to reflect general price changes, the supply curve is nearly vertical. At the other extreme when general prices are stable (σ^2 is relatively small) the slope of the supply curve approaches the limiting value of γ.[4]

B. Completion and Solution of the Model

A central assumption in the development above is that supply behavior is based on the *correct* distribution of the unobserved current price level, P_t. To proceed, then, it is necessary to determine what this correct distribution is, a step which requires the completion of the model by inclusion of an aggregate demand side.

As suggested earlier, this will be done by postulating a demand function for goods of the form:

$$y_t + P_t = x_t \tag{8}$$

where x_t is an exogenous shift variable—equal to the observable *log* of nominal *GNP*. Further, let $\{\Delta x_t\}$ be a sequence of independent, normal variates with mean δ and variance σ_x^2.[5]

[4] This predicted relationship between a supply elasticity and the variance of a component of the price series is analogous to the link between the income elasticity of consumption demand and the variances of permanent and transitory income components which Friedman (1957) observes. As will be seen in Section II, it works in empirical testing in much the same way as well.

[5] This particular characterization of the "shocks" to the economy is not central to the theory, but to discuss

The relevant history of the economy then consists (at most) of y_{nt} (which fixes calendar time), the demand shifts x_t, $x_{t-1}, \ldots$, and past actual real outputs $y_{t-1}, y_{t-2}, \ldots$. Since the model is linear in *logs*, it is reasonable to conjecture a price solution of the form:[6]

$$P_t = \pi_0 + \pi_1 x_t + \pi_2 x_{t-1} + \pi_3 x_{t-2} + \ldots + \eta_1 y_{t-1} + \eta_2 y_{t-2} + \ldots + \xi_0 y_{nt} \tag{9}$$

Then $\overline{P}_t$ will be the expectation of P_t, based on all information *except* x_t (the current demand level) or:

$$\begin{aligned}\overline{P}_t = \overline{P}_0 &+ \pi_1(x_{t-1} + \delta) + \pi_2 x_{t-1} \\ &+ \pi_3 x_{t-2} + \ldots + \eta_1 y_{t-1} \\ &+ \eta_2 y_{t-2} + \ldots + \xi_0 y_{nt}\end{aligned} \tag{10}$$

To solve for the unknown parameters π_i, η_j and ξ_0 we first eliminate y_t between (7) and (8), or equate quantity demanded and supplied. Then inserting the right sides of (9) and (10) in place of P_t and $\overline{P}_t$, one obtains an identity in $\{x_t\}$, $\{y_t\}$, and y_{nt}, which is then used to obtain the parameter values. The resulting solutions for price and output are:[7]

$$P_t = \frac{\theta\gamma\delta}{1+\theta\gamma} - \lambda\beta + \frac{1}{1+\theta\gamma} x_t + \frac{\theta\gamma}{1+\theta\gamma} x_{t-1} - \lambda y_{t-1} - (1-\lambda) y_{nt}$$

$$y_t = -\frac{\theta\gamma\delta}{1+\theta\gamma} + \lambda\beta + \frac{\theta\gamma}{1+\theta\gamma}\Delta x_t + \lambda y_{t-1} + (1-\lambda) y_{nt}$$

In terms of ΔP_t and y_{ct}, and letting $\pi = \theta\gamma/(1+\theta\gamma)$, the solutions are:

$$y_{ct} = -\pi\delta + \pi\Delta x_t + \lambda y_{c,t-1} \tag{11}$$

$$\Delta P_t = -\beta + (1-\pi)\Delta x_t + \pi\Delta x_{t-1} - \lambda\Delta y_{c,t-1} \tag{12}$$

Let us review these solutions for internal consistency. Evidently, P_t is normally distributed about $\overline{P}_t$. The conditional variance of P_t will have the constant (as assumed) variance $1/(1+\theta\gamma)^2\sigma_x^2$. Thus those features of the behavior of prices which were assumed "known" by suppliers in subsection A are, in fact, true in this economy.

To review, equations (11) and (12) are the *equilibrium values* of the inflation rate and real output (as a percentage deviation from trend). They give the intersection points of an aggregate demand schedule, shifted by changes in x_t, and an aggregate supply schedule shifted by variables (lagged prices) which determine expectations. In order to avoid the introduction of an additional, spurious "expectations parameter," one cannot solve for this intersection on a period-by-period basis; accordingly, we have adopted a method which yields equilibrium "paths" of prices and output. Otherwise, the interpretation of (11) and (12) is entirely conventional.

Not surprisingly, the solution values of inflation and the cyclical component of real output are indicated by (11) and (12) to be distributed lags of current and past changes in nominal output. A change in the nominal expansion rate, Δx_t, has an immediate effect on real output, and lagged effects which decay geometrically. The

rational expectations formation at all, *some* explicit stochastic description is clearly required. Independence is used here partly for simplicity, partly because it is empirically roughly accurate for most countries in the sample. The effect of autocorrelation in the shocks would, as can be easily traced out, be to add higher order lag terms to the solutions found below.

[6] This solution method is adapted from Lucas (1972), which is in turn based on the ideas of John Muth.

[7] If a demand function of the form $y_t = \xi P_t + x_t$ had been used, these solutions would assume the same form, with different expressions for the coefficients. If $\xi \neq 1$, however, x_t is an unobserved shock, unequal in general to observed nominal income. In this case, the model still predicts the time-series structure (moments and lagged moments) of the series y_{ct} and ΔP_t and is thus, in principle, testable. I have found empirical experimenting along these lines suggestive, but the series used are simply too short to yield results of any reliability.

immediate effect on prices is one minus the real output effect, with the remainder of the impact coming in the succeeding period. We note in particular that this lag pattern may well produce periods of simultaneous inflation and below average real output. Though these periods arise because of supply shifts, the shifts result from lagged perception of demand changes, and *not* from autonomous changes in the cost structure of suppliers.

In addition to these features, the model does indeed assert the existence of a natural rate of output: the *average* rate of demand expansion, δ, appears in (11) with a coefficient equal in magnitude to the coefficient of the current rate, and with the opposite sign. Thus changes in the average rate of nominal income growth will have *no* effect on average real output. On the other hand, unanticipated demand shifts do have output effects, with magnitude given by the parameter π. Since this effect depends on "fooling" suppliers (in the sense of subsection A), one expects that π will be larger the smaller the variance of the demand shifts. We next develop this implication explicitly.

From the definition of π in terms of θ and γ, and the definition of θ in terms of σ^2 and τ^2 we have

$$\pi = \frac{\tau^2\gamma}{\sigma^2 + \tau^2(1+\gamma)}$$

Combining with the expression for σ^2 obtained above, this gives

$$\pi = \frac{\tau^2\gamma}{(1-\pi)^2\sigma_x^2 + \tau^2(1+\gamma)} \tag{13}$$

For fixed τ^2 and γ, then, π takes the value $\gamma/(1+\gamma)$ at $\sigma_x^2=0$ and tends monotonically to zero as σ_x^2 tends to infinity.

The prediction that the average deviation of output from trend, $E(y_{ct})$, is invariant under demand policies is not, of course, subject to test: the deviations from a *fitted* trend line must average to zero. Accordingly, we must base tests of the natural rate hypothesis (in this context) on (13): a relationship between an observable variance and a slope parameter.

II. Test Results

Testing the hypothesis advanced above involves two steps. First, within each country (11) and (12) should perform reasonably well. In particular, under the presumption that demand fluctuations are the major source of variation in ΔP_t and y_{ct}, the fits should be "good." The estimated values of π and λ should be between zero and one. Finally, since (11) and (12) involve five slope parameters but only two theoretical ones, the estimated π and λ values obtained from fitting (11) should work reasonably well in explaining variations in ΔP_t.

The main object of this study, however, is not to "explain" output and price level movements within a given country, but rather to see whether the terms of the output-inflation "tradeoff" vary across countries in the way predicted by the natural rate theory. For this purpose, we shall utilize the theoretical relationship (13) and the estimated values of π and σ_x^2. Under the assumption that τ^2 and γ are relatively stable across countries, the estimated π values should decline as the sample variance of Δx_t increases.

Descriptive statistics for the eighteen countries in the sample are given in Table 1.[8] As is evident, there is no association

[8] The raw data on real and nominal *GNP* are from *Yearbook of National Accounts Statistics*, where series from many countries are collected and put on a uniform basis. The choice of countries is by no means random: the eighteen used are all the countries from which continuous series are available. The sample could thus be broadened considerably by use of sources from individual countries. To obtain the variables used in the tests, the *logs* of real and nominal output, y_t and x_t, are *logs* of the series in the source. The *log* of the price level, P_t, is the difference $x_t - y_t$; y_{ct} is the residual from the trend line $y_t = a + bt$, fit by least squares from the sample

TABLE 1—DESCRIPTIVE STATISTICS, 1952–67

Country	Mean Δy_t	Mean ΔP_t	Variance y_{ct}	Variance ΔP_t	Variance Δx_t
Argentina	.026	.220	.00096	.01998	.01555
Austria	.048	.038	.00104	.00113	.00124
Belgium	.034	.021	.00075	.00033	.00072
Canada	.043	.024	.00109	.00018	.00139
Denmark	.039	.041	.00082	.00038	.00084
West Germany	.056	.026	.00147	.00026	.00073
Guatemala	.046	.004	.00111	.00079	.00096
Honduras	.044	.012	.00042	.00084	.00109
Ireland	.025	.038	.00139	.00060	.00111
Italy	.053	.032	.00022	.00044	.00040
Netherlands	.047	.036	.00055	.00043	.00101
Norway	.038	.034	.00092	.00033	.00098
Paraguay	.054	.157	.00488	.03192	.03450
Puerto Rico	.058	.024	.00205	.00021	.00077
Sweden	.039	.036	.00030	.00043	.00041
United Kingdom	.028	.034	.00022	.00037	.00014
United States	.036	.019	.00105	.00007	.00064
Venezuela	.060	.016	.00175	.00068	.00127

between average real growth rates and average rates of inflation: this fact seems to be consistent with both the conventional and natural rate views of the tradeoff. Since our interest is in comparing real output and price behavior under different time patterns of nominal income, these statistics are somewhat disappointing. Essentially two types of nominal income behavior are observed: the highly volatile and expansive policies of Argentina and Paraguay, and the relatively smooth and moderately expansive policies of the remaining sixteen countries. But if the sample provides only two "points," they are indeed widely separated: the estimated variance of demand in the high inflation countries is on the order of 10 times that in the stable price countries.

The first three columns of Table 2 summarize the performance of equation (11) in accounting for movements in y_{ct}. The estimated values for π all lie between zero and one; with the exceptions of Argentina and Puerto Rico, so do the estimated λ values. The R^2s indicate that for many, or perhaps most countries, important output-determining variables have been omitted from the model. The R^2s for the inflation rate equation, (12), are given in column (4) of Table 2. In general, these tend to be lower than for equation (11), and not surprisingly the estimated coefficients from (12) (which are not shown) tend to behave erratically. Column (5) of Table 2 gives the fraction of the variance of ΔP_t explained by (12) when the coefficient estimates from (11) are imposed. (A "−" indicates a negative value.)[9]

With respect to its performance as an intracountry model of income and price determination, then, the system (11)–(12) passes the formal tests of significance. On the other hand, the goodness-of-fit statis-

period. The moments given in Table 1 are maximum likelihood estimates based on these series. The estimates reported in Table 2 are by ordinary least squares.

[9] The loss of explanatory power when these coefficients are imposed on (12) can be assessed formally by an approximate Chi-square test. By this measure, the loss is significant at the .05 level for Paraguay only. As Table 2 shows, however, this test is somewhat deceptive: for several countries the least squares estimates of (12) are so poor that there is little explanatory power to lose, and the test is "passed" vacuously.

TABLE 2—SUMMARY STATISTICS BY COUNTRY, 1953–67

Country	π	λ	R_y^2	$R_{\Delta P}^2$	R_ω^2
Argentina	.011 (.070)	−.126 (.258)	.018	.929	.914
Austria	.319 (.179)	.703 (.209)	.507	.518	—
Belgium	.502 (.100)	.741 (.093)	.875	.772	.661
Canada	.759 (.064)	.736 (.075)	.936	.418	—
Denmark	.571 (.118)	.679 (.110)	.812	.498	.282
West Germany	.820 (.136)	.784 (.110)	.881	.130	—
Guatemala	.674 (.301)	.695 (.274)	.356	.016	—
Honduras	.287 (.152)	.414 (.250)	.274	.521	.358
Ireland	.430 (.121)	.858 (.111)	.847	.499	.192
Italy	.622 (.134)	.042 (.183)	.746	.934	.914
Netherlands	.531 (.111)	.571 (.149)	.711	.627	.580
Norway	.530 (.088)	.841 (.096)	.893	.633	.427
Paraguay	.022 (.079)	.742 (.201)	.568	.941	.751
Puerto Rico	.689 (.121)	1.029 (.072)	.939	.419	—
Sweden	.287 (.166)	.584 (.186)	.525	.648	.405
United Kingdom	.665 (.290)	.178 (.209)	.394	.266	.115
United States	.910 (.086)	.887 (.070)	.945	.571	.464
Venezuela	.514 (.183)	.937 (.148)	.755	.425	—

tics are generally considerably poorer than we have come to expect from annual time-series models.

In contrast to these somewhat mixed results, the behavior of the estimated π values across countries is in striking conformity with the natural rate hypothesis. For the sixteen stable price countries, $\hat{\pi}$ ranges from .287 to .910; for the two volatile price countries, this estimate is smaller by a factor of 10! To illustrate this order-of-magnitude effect more sharply, let us examine the complete results for two countries: the United States and Argentina. For the United States, the fitted versions of (11) and (12) are:

$$y_{ct} = -.049 + (.910)\Delta x_t + (.887)y_{c,t-1}$$

$$\Delta P_t = -.028 + (.119)\Delta x_t + (.758)\Delta x_{t-1} - (.637)\Delta y_{c,t-1}$$

The comparable results for Argentina are:

$$y_{ct} = -.006 + (.011)\Delta x_t - (.126)y_{c,t-1}$$

$$\Delta P_t = -.047 + (1.140)\Delta x_t - (.083)\Delta x_{t-1} + (.102)\Delta y_{c,t-1}$$

In a stable price country like the United States, then, policies which increase nomi-

nal income tend to have a large initial effect on real output, together with a small, positive initial effect on the rate of inflation. Thus the apparent short-term tradeoff is favorable, as long as it remains unused. In contrast, in a volatile price country like Argentina, nominal income changes are associated with equal, contemporaneous price movements with no discernible effect on real output. These results are, of course, inconsistent with the existence of even moderately stable Phillips curves. On the other hand, they follow directly from the view that inflation stimulates real output if, and only if, it succeeds in "fooling" suppliers of labor and goods into thinking *relative* prices are moving in their favor.

III. Concluding Remarks

The basic idea underlying the tests reported above is extremely simple, yet I am afraid it may have become obscured by the rather special model in which it is embodied. In this section, I shall try to restate this idea in a way which, though not quite accurate enough to form the basis for econometric work, conveys its essential feature more directly.

The propositions to be compared empirically refer to the effects of aggregate demand policies which tend to move inflation rates and output (relative to trend) in the same direction, or alternatively, unemployment and inflation in opposite directions. The conventional Phillips curve account of this observed co-movement says that the terms of the tradeoff arise from relatively stable structural features of the economy, and are thus independent of the nature of the aggregate demand policy pursued. The alternative explanation of the same observed tradeoff is that the positive association of price changes and output arises because suppliers misinterpret general price movements for relative price changes. It follows from this view, first, that changes in average inflation rates will not increase average output, and secondly, that the higher the *variance* in average prices, the less "favorable" will be the observed tradeoff.

The most natural cross-national comparison of these propositions would seem to be a direct examination of the association of average inflation rates and average output, relative to "normal" or "full employment." Unfortunately, there seems to be no satisfactory way to measure normal output. The deviation-from-fitted-trend method I have used *defines* normal output to be average output. The use of unemployment series suffers from the same difficulty, since one must somehow select the (obviously positive) rate to be denoted full employment.

Thus although the issue revolves around the relation between *means* of inflation and output rates, it cannot be resolved by examination of sample averages. Fortunately, the existence of a stable tradeoff also implies a relationship between *variances* of inflation and output rates, as illustrated in Figure 1. With a stable tradeoff, policies which lead to wide variation in prices must also induce comparable variation in real output. If these sample variances do not tend to move together (and, as Table 1 shows, they do not) one

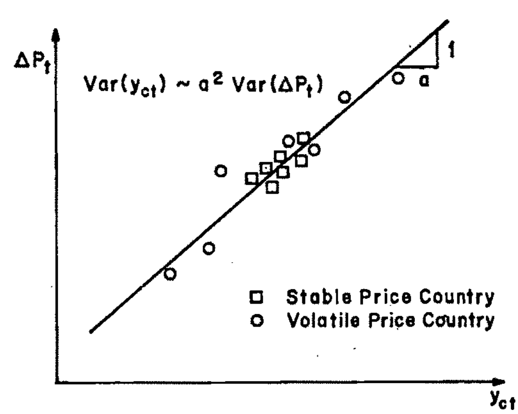


FIGURE 1

can only conclude that the tradeoff tends to fade away the more frequently it is used, or abused.

This simple argument leads to a formal test if the output-inflation association is entirely contemporaneous. In fact, however, it involves lagged effects which make a direct comparison of variances, as just suggested, difficult in short time-series. Accordingly, it has been necessary to impose a specific, simple structure on the data. As we have seen, this structure accounts for output and inflation rate movements only moderately well, but well enough to capture the main phenomenon predicted by the natural rate theory: the higher the variance of demand, the more unfavorable are the terms of the Phillips tradeoff.

REFERENCES

M. Friedman, *A Theory of the Consumption Function,* Princeton 1957.

———, "The Role of Monetary Policy," *Amer. Econ. Rev.*, Mar. 1968, *58,* 1–17.

D. F. Gordon and A. Hynes, "On the Theory of Price Dynamics," in E. S. Phelps et al., *Micro-economics of Inflation and Employment Theory,* New York 1969.

R. E. Lucas, Jr., "Expectations and the Neutrality of Money," *J. Econ. Theor.*, Apr. 1972, *4,* 103–24.

———, "Econometric Testing of the Natural Rate Hypothesis," *Conference on the Econometrics of Price Determination,* Washington 1972, 50–59.

——— and L. A. Rapping, "Real Wages, Employment and the Price Level," *J. Polit. Econ.*, Sept./Oct. 1969, *77,* 721–54.

——— and ———, "Unemployment in the Great Depression: Is There a Full Explanation?", *J. Polit. Econ.*, Jan./Feb. 1972, *80,* 186–91.

J. F. Muth, "Rational Expectations and the Theory of Price Movements," *Econometrica,* July 1961, *29,* 315–35.

E. S. Phelps, introductory chapter in E. S. Phelps et al., *Micro-economics of Inflation and Employment Theory,* New York 1969.

F. Raines, "Macroeconomic Demand and Supply: an Integrative Approach," Washington Univ. working paper, Apr. 1971.

A. Rees, "On Equilibrium in Labor Markets," *J. Polit. Econ.*, Mar./Apr. 1970, *78,* 306–10.

P. Taubman and M. Wilkinson, "User Cost, Capital Utilization and Investment Theory," *Int. Econ. Rev.*, June 1970, *11,* 209–15.

United Nations, Department of Economic and Social Affairs, United Nations Statistical Office, *Yearbook of National Accounts Statistics, 66 and 68,* New York 1958.

A Note on the Concept and Measure of Consumer's Surplus

By Michael E. Burns*

In an important new book, E. J. Mishan is critical of statements made by prominent economists that have thrown doubt on the usefulness of the concept of consumer's surplus. Mishan goes on to draw attention to the need for clarification of ". . . the concept itself, and the relationship between the concept and its measurable proxy, the demand curve" (1971, pp. 325–28). A fundamental obstacle to such clarification would appear to be a view quite generally held by economists, and formally stated by Paul Samuelson that ". . . the Marshallian concept of consumer's surplus does not refer to any one thing, but to at least half a dozen interrelated expressions."[1] Accordingly the approach offered here begins with a reinterpretation of the literature normally associated with the origins of the concept in question.

There is general acknowledgement that J. Dupuit gave the first description (1844) of the notion that Alfred Marshall later referred to in 1920 as consumer's surplus. What is generally forgotten is that Dupuit's primary concern was the *monetary* evaluation of the benefits (or losses) associated with alternative price policies for a public good.[2] His famous but perhaps slightly ambiguous conclusion may be illustrated using the toll-bridge example; the consumer's gain in *utility* brought about by the reduction in a toll should be measured in *money* terms by the area to the left of the demand curve (for the use of the bridge), between the old and new toll (price) levels.[3] His justification for the accuracy of this measure followed from two considerations. Firstly, his attempts at rationalization were influenced by considerations of services of an indivisible nature and led him to the plausible view that the monetary evaluation of the additional utility obtained from the purchase of a good should be taken as ". . . that maximum sacrifice (in income) that each consumer would be prepared to make, in order to be able to make the purchase of the good."[4] Secondly, Dupuit believed that the magnitude of this sacrifice was exactly measured in terms of the areas to the left of the demand curve. Any faults in the justification of the suggested evaluation of the benefits concerned should not conceal the unique nature of the concept whose measure was being sought.

Marshall's contribution followed very similar lines, but with the important qualification that the accuracy of the measure 'what a consumer would pay, rather than go without' required that the marginal utility of income be constant.[5]

* I have benefitted from discussions with colleagues at The University of Western Australia, and from the valuable comments of the referees on an earlier draft of this paper. Needless to say, all remaining errors are my own responsibility.

[1] See Samuelson (1947), pp. 197–202. Samuelson's discussions here are based on an earlier paper (1942), where he considered in detail the question of the constancy of the marginal utility of income.

[2] See Dupuit, in *Readings in Welfare Economics*, pp. 253–83.

[3] See Dupuit, in *Readings in Welfare Economics*, particularly pp. 270–71.

[4] In the original version, Dupuit wrote "L'économie politique doit prendre pour mesure de l'utilité d'un objet le sacrifice maximum que chaque consommateur serait disposé à faire pour se le procurer" (p. 40).

[5] See Marshall, p. 842. There is often some confusion between the marginal utility of income and that of money. In the text we are always referring to the former concept. Don Patinkin, (1956), p. 65, makes the distinc-

However it is interesting to reconsider Marshall's introduction of the now familiar terminology, substituting the expression *gain in utility* for what Marshall referred to as a surplus of satisfaction:

> . . . [the consumer] derives from a purchase a surplus of satisfaction [gain in utility]. The excess of the price which he would be willing to pay rather than go without the thing, over that which he actually does pay, is the economic [money] measure of this surplus of satisfaction [gain in utility]. It may be called consumer's surplus. [p. 124]

It is perhaps signficant that J. R. Hicks, without explicitly making the substitution of terms suggested above, concluded that ". . . the best way of looking at consumer's surplus is to regard it as a means of expressing, in terms of money income, the gain which accrues to the consumer as a result of a price fall" (1947, p. 40).

The substantial agreement amongst these authors as to the precise nature of the concept they sought to measure provides a basis for the treatment given in this paper. Accordingly, in Section I it is shown that for an individual the "income equivalent" of a utility change effected by a *ceteris paribus* price change is in exact agreement with Dupuit's original measure. In Section II, I reconsider two of the alternative expressions that have been offered as measures of the income equivalent of such a utility change. Fresh insight is given into the approximations that are implicit in these expressions. Finally, in Section III, the basic results are extended in a number of directions. Questions concerning the market demand curve, the meaning of areas to the left of the demand curve and the possibility of evaluating utility changes effected by changes in more than one variable are all examined in some detail.

tion between the two terms quite clear, as does Samuelson (1947), p. 190.

I. The Income Equivalent of a Utility Change

Before we undertake the formal derivation, some insight can be obtained by considering a simple illustrative example. Suppose quite generally that some of the conditions facing a utility maximizing consumer are altered, so that the consumer moves to a new equilibrium position, undergoing a change in the level of his total utility. Suppose further that the changes in each of the variables concerned are sufficiently small for the marginal utility of income to be treated as constant throughout the movement. Given any such change in utility, and the relevant value of the marginal utility of income, there must exist a sum of money (an income equivalent) that satisfies the relationship

$$\begin{aligned}&\text{(Change in Utility)}\\ &\equiv \text{(marginal utility of income)}\\ &\times \text{(income equivalent)}\end{aligned} \tag{1}$$

Given that the marginal utility of income cannot be constant everywhere, we may thus note that equal utility changes between different pairs of situations would generally be expected to yield different income equivalents.

Let us now consider the position of a utility-maximizing consumer, whose (ordinal) utility function is described by

$$U = f(x_1 \ldots x_n) \tag{2}$$

the properties of the function being such as enable any small change in utility to be stated as

$$dU = \sum_{i=1}^{n} \frac{\delta f}{\delta x_i} dx_i \tag{3}$$

Before any changes take place, the revenue apportioned to the ith good is

$$R_i = p_i x_i \qquad (i = 1 \ldots n) \tag{4}$$

so that any small change in this apportion-

ment may be described by

$$(5)\quad dR_i = p_i dx_i + x_i dp_i \qquad (i = 1 \ldots n)$$

Rewriting (5) in terms of dx_i and substituting into (3) yields

$$(6)\quad dU = \sum_{i=1}^{n} \left(\frac{\delta f}{\delta x_i} \bigg/ p_i \right) (dR_i - x_i dp_i)$$

However, the condition of utility maximization requires that

$$(7)\quad \frac{\delta f}{\delta x_i} \bigg/ p_i = \frac{\delta f}{\delta x_j} \bigg/ p_j = \lambda$$

$$= (\text{marginal utility of income})$$

$$(i, j = 1 \ldots n)$$

so that (6) reduces to

$$(8)\quad dU = \lambda \sum_{i=1}^{n} (dR_i - x_i dp_i)$$

$$= \lambda \left(dY - \sum_{i=1}^{n} x_i dp_i \right)$$

By comparing (8) with the considerations that led to (1), it is clear that the bracketed term $(dY - \sum_{i=1}^{n} x_i dp_i)$ can only be the exact income equivalent of a small utility change, as effected by a completely general set of changes in the parameters facing the consumer. Before we leave the general case for the time being, we should note a powerful result concerning changes that do not affect the individual's utility level. Since λ will not normally take the value zero, the statement

$$(9)\quad dU = 0 = \left(dY - \sum_{i=1}^{n} x_i dp_i \right)$$

provides concise information concerning movements along an indifference curve.

If the circumstances that bring about the utility change are of a restricted form, further simplifications of (8) are possible. Thus, if only income changes, all prices remaining constant, it follows from (8) that the income equivalent of the resulting change in utility is simply the change in income itself. Turning to utility changes brought about within the particular framework of a *ceteris paribus* change in the price of the ith good, we see that (8) reduces to

$$(10)\qquad dU = \lambda(-x_i dp_i)$$

Thus the income equivalent in this instance is the area to the left of the demand curve between the old and new price levels, a measure in agreement with that originally suggested by Dupuit. As we would expect, (10) tells us that an increase in price yields a negative income equivalent whereas a price fall provides a positive income equivalent consistent with a gain in utility. We may note that in the trivial case of a constant marginal utility of income, there is a direct numerical relationship between utility and areas to the left of the demand curve.

The question remains as to whether we may meaningfully evaluate the income equivalent of a large change in utility brought about by a greater price change than the one previously considered. Had the change in utility been brought about solely by a large change in income, no conceptual difficulty would arise from treating this income change as the sum of a series of continuous small changes. Bearing in mind that the income equivalent of a utility change effected by a small change in income is that change in income itself, we see that the sum of the income equivalents provides a monetary evaluation of benefits received that is in agreement with the actual change in income experienced. This would suggest that in so far as it is useful to talk of an individual being (say) N dollars better off as the result of a large change in income, so may we sum the income equivalents of utility changes effected by a series of continuous small price changes and conclude, if such a sum should also equal N dollars, that the individual is

N dollars better off as a result of the total price change.

On the basis that we regard the income equivalent of a large *ceteris paribus* price change as the sum of the income equivalents obtained by considering continuous small price changes, we may note that this concept is measured exactly by the area to the left of the demand curve, between the old and new price levels.

II. The Nature of the Compensating and Equivalent Variations

As stated earlier, a number of alternative expressions have been offered as measures of the 'income equivalent' of a utility change effected by a *ceteris paribus* price change. We shall consider two particular proxies that have received a good deal of attention in the literature, but at the same time we may note that any of the other proxies suggested could be submitted to a similar treatment. In both cases we shall limit ourselves to working through the result of a price fall for a normal good, although the expressions themselves apply equally to a price increase or decrease for any type of good.

Hicks originally appeared to believe that the desired income equivalent was equal to the sum of money (income) to be paid by a consumer when a price falls (received by him when a price rises), that would leave his general level of welfare unchanged.[6] Hicks called this the *compensating variation* definition of consumer's surplus, but following a comment by Henderson he recognized the approximation implicit in this measure and agreed that some account should be taken of an alternative measure, one that he chose to call the *equivalent variation* definition of consumer's surplus. For this concept we ask, what income would a consumer have to receive following a price fall (forego following a price rise) in order to be as well off at the old prices as he would be with his initial income at the new prices.

We may illustrate both of these concepts in Figure 1. Quantities of the ith good are measured horizontally along OX_i, while income (all other goods at constant prices) is measured on the vertical axis OY. With a fixed income Y_0 the budget constraint is given initially by Y_0P_1, and after the fall in the price of the ith good, by Y_0P_2. The indifference curves I_1 and I_2 show the utility levels that are optimal and feasible at the old and new price levels, respectively.

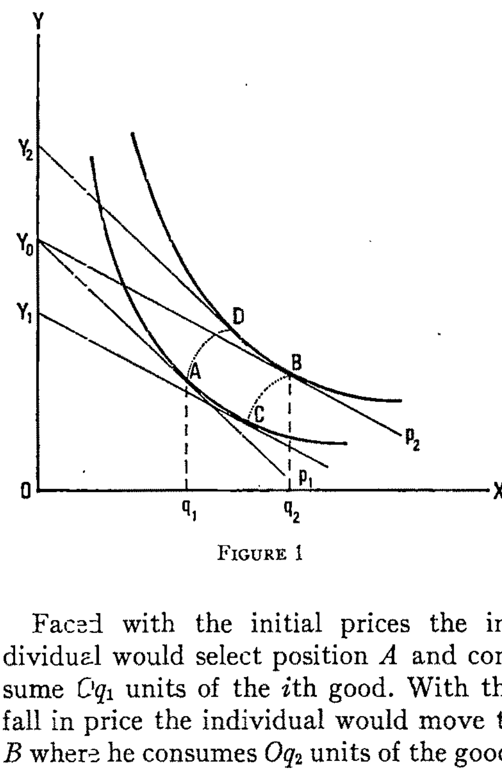


FIGURE 1

Faced with the initial prices the individual would select position A and consume Oq_1 units of the ith good. With the fall in price the individual would move to B where he consumes Oq_2 units of the good. What is required is the income equivalent of the utility change involved in the movement from A to B brought about by the price change. First we notice that the consumer could forego (Y_0-Y_1) of his income and still be as well off with the new

[6] See Hicks, pp. 40–41. The whole discussion in this paragraph should also be related to what Hicks has to say later, pp. 330–32.

prices as he was in the initial situation. The sum (Y_0-Y_1) thus provides the compensating variation proxy that we described above. Then again we can see that the consumer would need to have a total income of Y_2 to be as well off with the old prices as he would have been with the initial income and the new prices. The sum (Y_2-Y_0) is the equivalent variation measure that Hicks proposed. Both these expressions clearly describe an income equivalent of the utility difference between the indifference curves I_1 and I_2. Equally clearly the expressions are generally unequal, and a very brief consideration will show that neither expression describes the income equivalent of the utility change due to the price fall.

The key to the discrepancy is simply that any evaluation of the income equivalent of a given utility change will depend upon the value of the marginal utility of income along the path on which the evaluation takes place. Our assumption that the ith good is normal yields precise information as to the relevant values of the marginal utility of income, in particular that it will be greater at C than at either A or B whereas it must be less at D than at either of these terminal positions.[7] Thus the compensating variation will obtain the income equivalent of the utility difference between I_1 and I_2 taking into account the relatively high values of the marginal utility of income along the path (dotted) CB, while at the other extreme the equivalent variation will obtain an accordingly greater income equivalent of the same utility difference because of the lower values of the marginal utility of income of the path (dotted) AD.

Thus we can straightforwardly see why the expressions considered will generally take different values, and can only be approximations to the income equivalent of a utility gain associated with a price fall, which would be evaluated over the intermediate path AB. The magnitude of the differences between the compensating and the equivalent variations can easily be determined. Referring again to Figure 1, since the level of income is the same at both initial and final situations the sum of the income adjustments must equal zero along any path of adjustment between A and B that we may care to consider. Suppose we had undergone a series of price and income changes that first took us along the indifference curve I_1 from A to C, and then adjusted from C to B by income changes only. From equation (9) we see that the magnitude of the income changes along I_1 must be given by the integral $\int x_i dp_i$, evaluated over AC. This magnitude must equal that of the sum of the income adjustments from C to B, which is of course the compensating variation. A similar argument can be used to show that the equivalent variation will take the value of the integral $\int x_i dp_i$, evaluated along the indifference curve I_2 over the path DB. Since the price changes are the same in both integrals, it is the values of the x_i over the paths considered that determine relative magnitudes. Thus for vertically parallel indifference curves the expressions are equal. It also follows from these considerations that the difference between the expressions vanishes as the size of the total price change approaches zero.

III. Some Extensions of the Analysis

We have so far paid particular attention to the result that, for an individual, the income equivalent of a utility change effected by a *ceteris paribus* price change may be exactly measured by the area to the left of the relevant demand curve, between the initial and final price levels.

[7] To obtain these results it is only needed to notice that income is defined as "all other goods at constant prices," and that the marginal utility of income is given by the marginal utility of any good divided by its price.

This is a useful result, and amongst other things enables an appropriate interpretation to be made of the "Marshallian Triangle" that has received so much attention in the literature. Specifically we note that such areas do not refer to a utility gain achieved solely through the additional consumption of a specific good, and further that the *ceteris paribus* condition precludes the straightforward addition of such areas associated with different goods. However, treatment of the effects of a single price change on an individual represent a very limited application of our earlier analysis, and some more general results can easily be obtained. As we stated above, the result

$$dU = \lambda\left(dY - \sum_{i=1}^{n} x_i dp_i\right) \tag{8}$$

enables us to determine, for an individual, the income equivalent of a utility change associated with a completely general set of simultaneous small changes in prices and income. Again, it is clear from the form of (8), that market observations of income, price, and quantity will straightforwardly yield the net sum of the income equivalent of each individual's utility differences effected by any such changes as we may care to consider.[8] Noting that these aggregate measures may always be obtained, it is instructive to continue for the time being with analysis at the individual level.[9]

Returning to the general result (8), caution must be exercised in the treatment of greater than incremental changes. For example, suppose that we wish to compare an initial situation A and a final situation B characterized by substantially different price and income levels. It follows immediately from consideration of (8) that we have a necessary and sufficient condition for B to be preferred to A. This is that there should exist at least one path of small continuous price and income adjustments from A to B that yields a nonnegative sequence of incremental income equivalents, at least one element being positive. However our primary concern is to obtain a useful monetary evaluation of the utility difference between alternative situations, and the question we need to consider when we are dealing with large multiple price and income changes is how far we may still regard such an evaluation as being measured by the sum of successive incremental income equivalents. The complication that arises in such a case is that by allowing more than one variable to change we are admitting the possibility of an infinite number of adjustment paths between any two terminal situations. Since the marginal utility of income may take quite a different range of values along each different adjustment path, so will the sum of the income equivalents take on different magnitudes depending upon the precise path taken between the situations concerned. The unique determination of the income equivalent of a utility change effected by a large *ceteris paribus* price change is thus possible simply because such a framework defines a unique adjustment path.[10]

This path-dependence was first considered by H. Hotelling in 1938 and is the subject of a more recent and very thorough study by Eugene Silberberg (1972).[11] It

[8] There is broad agreement between this paper and a recent article by Arnold Harberger, who notes a similar result (1971), p. 789. However, the general approaches adopted are quite different.

[9] There are certain problems relating to the use of aggregate measures, and whenever we refer to such measures the reader should bear in mind certain necessary qualifications. For a discussion of these matters, see David Winch (1965), (particularly pp. 406–07).

[10] The reader should note, following equation (9), that this "path-dependence" only applies when the level of utility changes. Any movements along an indifference curve or surface must have a zero income equivalent, although if one leaves an indifference curve, and then returns to it, there is of course no reason why the same result should hold.

[11] Had more attention been paid to Hotelling's dis-

requires that we recognize the impossibility of associating a unique evaluation in money terms with the utility difference between any two situations. This is not to say that some evaluations of the income equivalent of a particular utility change are not a great deal more interesting than others. Indeed, with reference to Figure 1, consider any adjustments that would take us along the indifference curve I_1 from A to C and then from C to B solely as a result of income changes. The income equivalent evaluated along such a path would exactly equal Hicks' *compensating variation.*[12] An alternative path (AD by income changes, DB along I_1) would yield an income equivalent equal to the *equivalent variation*. Knowledge of the magnitude of these expressions could be useful for at least three reasons:

(i) for an individual, the sign of both these measures must be in agreement with the sign of the utility change concerned;

(ii) at the aggregate level, the sum of either of the individual *compensating* or *equivalent variations* will yield information as to whether compensations could be paid that would enable at least one individual to be made better off while leaving nobody worse off, although such information will sometimes be ambiguous;[13]

(iii) the Hicksian expressions can be regarded as providing sensible upper and lower limiting values of a monetary evaluation as the (net) benefits or losses involved.[14]

Consider now an income equivalent taken over a path of adjustment such that its value obtained is intermediate between the compensating and equivalent variations. The first criterion of usefulness is straightforwardly satisfied. In the second case our intermediate income equivalent would not be significantly different from zero in the ambiguous case, but could otherwise be generally expected to provide accurate information as to whether the stated compensations were possible. In the final case, the usefulness of having upper and lower limits would be related to the magnitude of the difference between the limits. In the substantial majority of cases where the marginal utility of income will not vary greatly over the range of the changes considered, the difference will be small and accordingly the intermediate income equivalent may be treated as a useful approximation to the desired monetary evaluation.

It remains to be noted that although neither of the Hicksian measures can be determined with any certainty in the aggregate form, it is a straightforward matter to determine the value of the (aggregate) income equivalent of a utility change evaluated along any adjustment path that might be observed or predicted to take place. Given the further point that for each individual most "actual" adjustment paths will be more direct than those that would yield the compensating or equivalent variations, we have every reason to expect that actual measures of the aggregate income equivalent would generally take values intermediate between the desired limiting expressions.

cussion of the matter (see particularly pp. 246–47), much of the confusion in the later literature might have been avoided. My own work was originally prepared in ignorance of Silberberg's paper and, although our approaches are quite different, certain similar conclusions are drawn.

[12] The adjustments along an indifference curve, yielding a zero utility change, must have a zero income equivalent. This also follows from result (9).

[13] It is quite possible for the sums of the individual compensating and equivalent variations to take different signs, given that there are both price falls and price rises.

[14] The two expressions may be regarded as evaluations of a fixed utility difference, taking into account the different (utility) purchasing power of money at the initial and final situations.

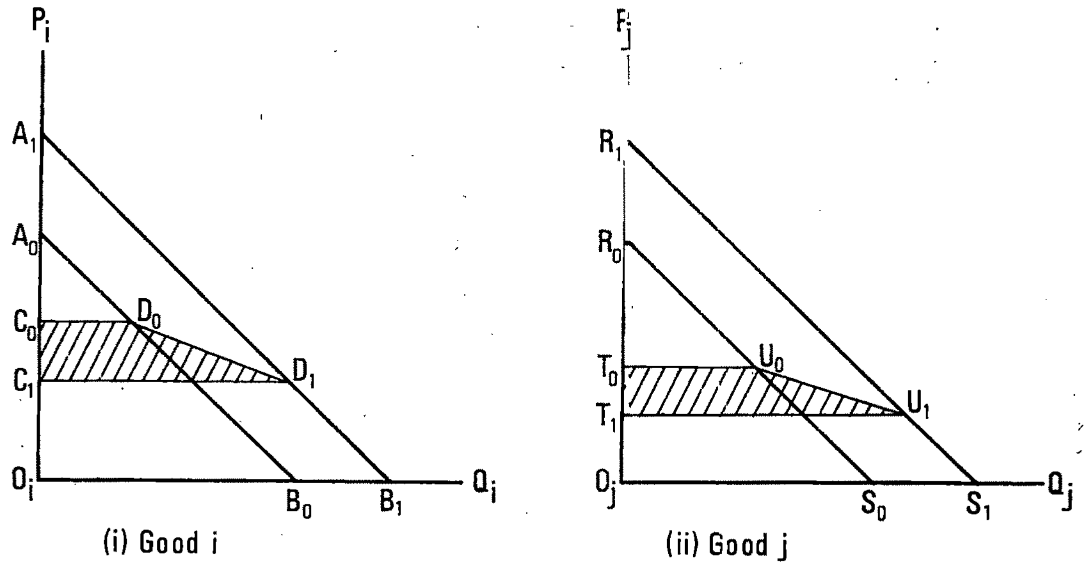


FIGURE 2

Having established reasons for our interest in the income equivalent of a utility change effected by a large change in more than one variable, it is only left for us to demonstrate the application of the analysis with a specific example. Suppose that the prices of the ith and jth goods change, income and all other prices held constant, so that (8) may be written

$$dU = \lambda(-x_i dp_i - x_j dp_j) \tag{11}$$

We illustrate the possible effects of a series of price changes of this type in Figure 2. Let A_0B_0 and R_0S_0 be the demand curves for the ith and jth goods, respectively, given the initial prices O_iC_0 and O_jT_0. If the prices change to O_iC_1 and O_jT_1, both demand curves will shift and we have taken the case where they both move to the right, A_0B_0 to A_1B_1 and R_0S_0 to R_1S_1. If we view these movements as the sum of many small movements along linear paths between D_0 and D_1 and between U_0 and U_1, respectively, the income equivalent of the change in total utility effected by these price changes is given by the sum of the two shaded areas $C_0C_1D_1D_0$ and $T_0T_1U_1U_0$. We can immediately see that for most practical purposes the variability of the income equivalent according to the path chosen is not going to be of any significant dimension. Indeed the normal processes of data observation or of prediction of the relevant price and quantity values will usually incorporate measurement errors far greater than the difference between the evaluations of the income equivalent taken along separate paths of price adjustment.[15]

Noting that path-dependence is usually of greater theoretical than practical importance, it is useful to consider briefly the evaluation of the income equivalent effected by changes that include a change in income itself. If income and the price of only the ith good are allowed to vary, (8) yields

$$dU = \lambda(dY - x_i dp_i) \tag{12}$$

which states that the income equivalent is the resulting sum of the change in income

[15] Most evaluations will lie between the Hicksian expressions, the difference between these latter terms being approximately $\Sigma dxdp$ for small changes.

itself, and an area that may be obtained by a treatment analogous to that used in the previous example. However, our interest in this case concerns the inclusion of the income change term. Suppose that instead of using the traditional model of consumer behavior where income is treated as an exogenous variable, we wished to allow an individual to substitute income for leisure. We might allow for this by writing the individual's utility function as

$$U = g(x_1 \ldots x_n, L) \tag{13}$$

and by defining income as

$$Y = (A - L)w \tag{14}$$

where A is the maximum possible time that may be allocated to the pursuit of income within the period considered, and w is the income forgone for an additional unit of leisure. To simplify the analysis, w as well as A may be treated as a constant. The reader may easily confirm that using this formulation the general result, equivalent of (8), is

$$dU = \lambda\left(\sum_{i=1}^{n} - x_i dp_i\right) \tag{15}$$

providing that income can only be changed by planned substitution between income and leisure. This result suggests that we should if possible distinguish between what we might call planned and unplanned variations in income, and only incorporate income changes of the second type into the evaluation of the income equivalent expression. As with the original model, no difficulty is presented in obtaining aggregate results, although the earlier cautionary comments still apply.

IV. Concluding Comments

In concluding our discussions of some of the applications and limitations of analysis based on the concept of consumer's surplus, it is appropriate to remind the reader of two points. First, for reasons stated earlier, we have dwelt very little on some of the problems associated with the aggregation of our results. Secondly, although we have talked of aggregate benefits or losses, these have only been considered in respect of consumption behavior and have ignored the shadow concept of producer's surplus.[16] We have done this to avoid repeating much that has appeared in the literature before, but our position concerning the second point should be quite clear. Any analysis of the type presented here, if used to evaluate benefits or losses accruing to the community as a whole, must incorporate the appropriate producer's surplus effects, and this latter concept should be treated with similar caution to that exercised in respect of our treatment of consumer's surplus.[17]

Notwithstanding these comments we would summarize the strong conclusions that can be drawn from this paper. The income equivalent of a utility change effected either by a *ceteris paribus* price change, or by small multiple price and income changes, is an entirely valid construct and may be measured accurately from observations of price, income, and quantity. The income equivalent of a utility change effected by large multiple price and income changes is a useful construct, and may also be measured accurately from observable data. Providing economists understand the concepts involved and the limitations and restrictions

[16] The recent 1971 survey article by J. M. Currie et al. provides a comprehensive summary of the relevant literature.

[17] The inclusion of producer's surplus effects may well reduce the importance of path-dependency, since its influence will often be counteracted between the two "surplus" measures. To illustrate this, consider the imposition of a sales tax on a particular good. Alternative adjustment paths can be generated by considering simultaneous variations in income, noting that an increase in income will generally shift the demand curve to the right and the supply curve to the left.

implicit in their use, the notions of consumer's and producer's surplus have a wide range of useful and illuminating applications in the fields of economic analysis and policy.

REFERENCES

J. M. Currie, J. A. Murphy and A. Schmitz, "The Concept of Economic Surplus and its Use in Economic Analysis," *Econ. J.*, Dec. 1971, *81*, 741–99.

J. Dupuit, "On the Measurement of the Utility of Public Works," translated in *International Economic Papers*, 2 (1952), 83–110; reprinted in *Readings in Welfare Economics*, Homewood 1969; "De l'utilitie et de sa Mesure," first published in 1844, reprinted with comments by M. de Bernardi and L. Einaudi, Turin and Paris, 1934.

A. C. Harberger, "Three Basic Postulates for Applied Welfare Economics: An Interpretive Essay," *J. Econ. Lit.*, Sept. 1971, *9*, 785–97.

A. Henderson, "Consumer's Surplus and the Compensation Variation," *Rev. Econ. Stud.*, 1941, *8*, 117–21.

J. R. Hicks, *Value and Capital*, 2d ed., Oxford 1947.

H. Hotelling, "The General Welfare in Relation to Problems of Taxation and of Railway and Utility Rates," *Econometrica*, July 1938, *6*, 242–69.

A. Marshall, *Principles of Economics*, 8th ed., London 1920.

E. J. Mishan, *Cost-Benefit Analysis*, London 1971.

D. Patinkin, *Money, Interest and Prices*, Evanston 1956.

P. A. Samuelson, "The Constancy of the Marginal Utility of Income," in O. Lange et al., eds., *Studies in Mathematical Economics and Econometrics in Memory of Henry Schultz*, Chicago 1942; also reprinted in *The Collected Scientific Papers of Paul A. Samuelson*, Cambridge 1965, ch. 5.

———, *Foundations of Economic Analysis*, Cambridge 1947.

E. Silberberg, "Duality and the Many Consumer's Surpluses," *Amer. Econ. Rev.*, Dec. 1972, *62*, 942–52.

D. M. Winch, "Consumer's Surplus and the Compensation Principle," *Amer. Econ. Rev.*, June 1965, *55*, 395–423.

Monetary and Fiscal Policy in a Two-Sector Aggregative Model

By DALE W. HENDERSON AND THOMAS J. SARGENT*

This paper analyzes the short-term influences exerted by monetary and fiscal policies in a two-sector macroeconomic model. The model is not new, but is identical with Hirofumi Uzawa's two-sector model, except that we have added markets in stocks of three paper assets, money, bonds, and equities. Adding these three markets produces essentially the same model used by Duncan Foley and Miguel Sidrauski to describe long-run behavior, given certain government policies. While the model studied here is virtually indistinguishable from theirs, we use the model to investigate very different questions. Our method of analysis is best characterized as "Keynesian" in two important senses. First, we analyze events over the "short run," a period of time so short that it is legitimate to assume that capital is fixed, even in the face of a nonzero rate of net investment. Where time is viewed as passing continuously, as it shall be here, this definition means that short-run occurrences are those that are assumed to take place at a point in time. Second, for the most part we assume that the money wage rate is rigid at any moment, the kind of assumption that it is necessary to make in order to build a model in which unemployment is possible, and in which it is interesting to investigate the short-run effects on employment and output of alternative monetary and fiscal policies. We assume only that the money wage rate is fixed at a point in time, which is compatible with having the money wage change over time, as long as it does so continuously. The money wage might well respond to the imposition of disturbances on the economy; our assumption permits such adjustments to occur over time, but rules out any instantaneous adjustments in money wages.

An important characteristic of the model is that it includes a perfect market in existing stocks of physical capital, a market in which there are no transactions costs either for buyers or sellers of capital.[1] By buying or selling capital in this market, individual firms are able to obtain instantaneously the quantities of capital they desire to hold at each moment. Monetary and fiscal policies work by introducing disturbances in this market, disturbances that ultimately affect the prices of capital goods and consumption goods, thereby affecting the number of workers firms want to employ at a fixed money wage. Thus, in our model, the market in existing stocks of physical capital occupies a critical role in transmitting monetary and fiscal policy influences. In contrast, the standard Keynesian model assumes no market in existing physical capital, presumably because of very high transactions costs. Such an assumption rationalizes the Keynesian flow investment demand function and makes the demand func-

* Member of the staff of the Board of Governors of the Federal Reserve System and member of the department of economics of the University of Minnesota, respectively. Sargent's work on this paper was partially supported by the Federal Reserve Bank of Minneapolis. The Federal Reserve System is in no way responsible for the conclusions of this paper.

[1] By transactions costs we mean a complete measure of the costs of transacting, including brokerage fees, bid-ask price spreads, search costs, dismantling and installation costs, etc.

tions for current output critical determinants of output.[2]

In our model, monetary policy is generally extremely potent in affecting output and employment. Of course, monetary policy also has important effects on these same variables under some sets of assumptions in the standard Keynesian model. However, in contrast with the Keynesian model, our model implies that an increase in the flow of government expenditures may either increase or reduce output and employment, depending upon the relative capital intensities of the two sectors and some other conditions. Moreover, whatever their direction, the effects of an increase in government expenditures are smaller the more similar are the capital intensities of the two sectors. Basically our model is an extension of James Tobin's Dynamic Aggregative Model, one that collapses to a version of Tobin's model when the production functions in the two sectors are assumed to be identical. When the production functions are identical, variations in government expenditures and in "autonomous" consumption have no effects on output and employment at a point in time, just as is true in Tobin's model. The "monetarist" implications of our model are ultimately attributable to the presence of the market in existing capital.

Section I contains a description of the behavior of firms, households, and the government. Section II describes the properties of the model, while our conclusions are presented in Section III.

[2] It seems clear that the formal model described by J. M. Keynes in the *General Theory* contains a market in existing stocks of capital goods. His discussions of the marginal efficiency of capital and the stock market (chs. 11 and 12) make this clear. The fact that the standard Keynesian model rules out trading in existing stocks of capital is perhaps the key departure that the Keynesians made from the formal model of the *General Theory*. For a discussion that attempts to highlight the important implications of what characteristics are assigned to the market in existing capital, see Sargent and Wallace.

I. Description of Firms, Households, and the Government

Firms

Firms in sector 1 use capital and labor to produce capital goods, those in sector 2 use the same kinds of capital and labor to produce consumer goods. Individual firms within each sector are faced with identical production functions. The price of capital goods is p_1p, while the price of consumer goods is p. Firms in both sectors are assumed to hire the amounts of capital and labor that maximize their profits at each moment. At each moment they can hire all the labor they want at the fixed money wage w, and they can purchase or sell all the capital they wish to at the price of capital, p_1p (alternatively, we might just as well assume that firms can rent all the capital they wish at an appropriate rental). Firms are assumed to issue no bonds, financing capital expansion by issuing equities.[3] However, we shall assume that when their yields are equal, government bonds and equities are perfect substitutes as far as wealth-holders are concerned, so that in equilibrium the nominal yield on equities must equal the nominal yield on government bonds, r. This assumption implies that r is the appropriate discount rate to be used in calculating the firm's present value; equivalently, r is the market interest rate that belongs in the cost of capital used to calculate the firm's profits.[4] Firms are assumed to retain no

[3] Our assumption that firms issue no bonds is not restrictive. For as Franco Modigliani and Merton Miller have shown, the pertinent cost of capital is the yield on equities for firms which have no bonds outstanding. Moreover, even if firms were permitted to have bonds outstanding, our assumption that wealth-holders regard bonds and equities as perfect substitutes would guarantee that the yield on a firm's equities would not depend on the volume of bonds it has outstanding. See Modigliani and Miller, in particular their "proposition II."

[4] See Dale Jorgenson for a demonstration of the equivalence of profit maximization and present value maximization, given a properly defined cost of capital.

earnings, paying out all profits in the form of dividends.

Outputs in sectors 1 and 2, Y_1 and Y_2, respectively, are determined by the production functions

$$Y_1 = Y_1(K_1, N_1) \tag{1}$$

$$Y_2 = Y_2(K_2, N_2) \tag{2}$$

where N_i is employment in the ith sector and K_i is capital in the ith sector. We assume that each production function is linearly homogeneous in K_i and N_i and has positive though diminishing marginal products, and positive cross-partial derivatives, i.e., $Y_{iN}>0$, $Y_{iK}>0$, $Y_{iNN}<0$, $Y_{iKK}<0$, $Y_{iKN}>0$, where Y_{iN} is the marginal product of labor in sector i, etc.

We assume that capital depreciates exponentially at the same rate δ in both sectors. Households and firms in both sectors expect that both the money wage, w, and the prices of consumer and capital goods will rise at the rate π, a magnitude that we treat as a parameter in the exercises below.[5] The government is assumed to levy a "profits" tax on each firm's net cash flow. The latter exceeds economic profits by the real yield on equities, $r-\pi$, multiplied by the value of the firm's capital. The marginal profits tax rate is t_K.

On these assumptions, the after-tax profits of the typical firm in sector 1, the capital goods sector, are[6]

$$\textstyle\sum_1 = pp_1Y_1(K_1, N_1) - wN_1 - (r+\delta-\pi)p_1pK_1 - t_K[p_1pY_1(K_1, N_1) - wN_1 - \delta p_1pK_1]$$

Rearranging the first-order conditions for profit maximization, we obtain

$$Y_{1N} = \frac{w}{p_1p} \tag{3}$$

$$Y_{1K} = \delta + \frac{1}{1-t_K}(r-\pi) \tag{4}$$

After-tax profits in sector 2 are

$$\textstyle\sum_2 = pY_2(K_2, N_2) - wN_2 - (r+\delta-\pi)p_1pK_2 - t_K[pY_2(K_2, N_2) - wN_2 - \delta p_1pK_2]$$

Upon rearranging the first-order conditions for profit maximization, we obtain

$$Y_{2N} = \frac{w}{p} \tag{5}$$

$$Y_{2K} = p_1\left[\delta + \frac{1}{1-t_K}(r-\pi)\right] \tag{6}$$

Equations (1) through (6) are assumed to hold at each moment in time, any violations of the equations being eliminated immediately by firms' trading in the markets for capital and labor.

Generally, we will take advantage of the linear homogeneity of (1) and (2) and write them in the intensive form

$$\frac{Y_i}{N_i} = Y_i\left(\frac{K_i}{N_i}, 1\right) = f_i(k_i)$$

where

$$k_i = \frac{K_i}{N_i}$$

In this notation, the marginal products of labor and capital are $Y_{iN}=f_i(k_i)-k_if_i'(k_i)$ and $Y_{iK}=f_i'(k_i)$. Substituting these expressions for the marginal products in equations (3) through (6) yields the form of the equations that it will be convenient to manipulate in Section II.

[5] It is necessary to treat π as a parameter at any moment in order to close the model. Otherwise the model would possess fewer equations than variables it must determine at a point in time. See Neil Wallace.

[6] It is easy to verify that maximizing profits so defined is equivalent to maximizing the firm's present value,

$$V(K_1, N_1, \dot{K}_1, t) = \int_0^\infty e^{-rt}\{(1-t_K)[p_1p(t)Y_1(K_1(t), N_1(t)) - w(t)N_1(t) - p_1p(t)\delta K_1(t)] - p_1p(t)\dot{K}_1(t)\}dt$$

where $w(t)$ and $p(t)$ are assumed to be expected to follow the paths

$$w(t) = we^{\pi t}$$
$$p(t) = pe^{\pi t}$$

The description of the factor markets is completed by noting that the total supply of capital, K, is fixed at a point in time. The assumptions that Y_{1K} and Y_{2K} are greater than zero, that there is a perfect market in existing capital, and that p_1, p, and r are flexible instantaneously imply that capital is fully employed at each moment,

$$K = K_1 + K_2 \tag{7}$$

However, the fact that the money wage is fixed means that labor need not be fully employed.

Notice that since households view bonds and equities as perfect substitutes, the value of equities issued by the capital goods industry, V_1, is the capitalized value of dividends, the appropriate capitalization rate being the bond rate, r, minus the anticipated rate of inflation:

$$V_1 = \frac{(1 - t_K)(pp_1 Y_1 - wN_1 - \delta p_1 p K_1)}{r - \pi}$$

which, by the first-order conditions for profit maximization and the linear homogeneity of the production function, equals $p_1 p\ K_1$. Similarly, the value of outstanding equities issued by the consumer goods industry, V_2, equals $p_1 p\ K_2$.

The Government

As liabilities the government has outstanding stocks of money with nominal value M and variable-interest-rate government bonds with nominal value B. The bonds are like savings deposits, their value remaining fixed in the face of fluctuations in the interest rate. The nominal yield on money is fixed at zero, while the nominal yield on bonds, r, is determined in the market. At a point in time the government conducts open-market operations subject to the constraint $dM = -dB$.

The government levies an income tax with marginal rate t_Y and a profits tax with marginal rate t_K. The base for the income tax is personal income, which consists of wage returns to labor, dividends, and nominal capital gains on equity. In real terms, laborers receive (w/p) $(N_1 + N_2)$, while stockholders receive in dividends

$$(1 - t_K)\left(p_1 Y_1 - \frac{w}{p} N_1 - \delta p_1 K_1\right) + (1 - t_K)\left(Y_2 - \frac{w}{p} N_2 - \delta p_1 K_2\right)$$

which, by the first-order conditions for profit maximization and the linear homogeneity of the production functions, equals $(r-\pi)p_1K$. Nominal capital gains on equities are expected to occur at the rate $\pi p_1 p K$. Thus, individuals expect to be paying income tax at the real rate

$$T_Y = t_Y\left[\frac{w}{p} N + (r - \pi)p_1 K + \pi p_1 K\right]$$

where $N = N_1 + N_2$. Notice that the after-tax real yield on equities equals $(1-t_Y)r-\pi$.

Real revenues from the profits tax are

$$T_K = t_K\left[p_1 Y_1 - \frac{w}{p} N_1 - \delta p_1 K_1\right] + t_K\left[Y_2 - \frac{w}{p} N_2 - \delta p_1 K_2\right]$$

which, by the first-order conditions for profit maximization and the linear homogeneity of Y_1 and Y_2, equals $(t_K/(1-t_K))$ $(r-\pi)p_1K$.

In real terms, total tax receipts net of transfer payments are given by $T = T_0 + T_Y + T_K$ where T_0 represents real tax receipts net of transfer payments that do not depend on personal income or business profits.[7]

Total real government expenditures

[7] Notice that there is a double tax on dividends.

denominated in units of consumer goods, G, are composed of real expenditures on capital goods, p_1g_1, and real expenditures on consumer goods, g_2. We will assume that if g_1 exceeds zero, the government simply stores or destroys the capital it is purchasing, neither renting it to private firms nor using it in production itself.[8]

The government makes its expenditure decisions subject to the flow budget constraint

$$G = T + \frac{DM}{p} + \frac{DB}{p}$$

where D is the operator d/dt. The government can finance a deficit by issuing money or bonds at some rate per unit time.

Households

Households own money, bonds, and equities. The real value of households' wealth is thus

$$W = \frac{M}{p} + \frac{B}{p} + p_1K$$

Portfolio balance for households is satisfied when the supply of real balances equals the demand,

$$(8) \quad \frac{M}{p} = m[Y, (1 - t_Y)r], m_1 > 0, m_2 \leq 0$$

where Y is real GNP, that is

$$(9) \quad Y = p_1Y_1 + Y_2$$

[8] We make this last assumption only for the sake of simplicity. Allowing the government to rent out the capital it buys to firms would require relatively minor changes in some definitions and accounting relationships. The rate of accumulation of capital to be used in production would be higher with government rental of capital. However, none of the conclusions of the comparative statics analysis presented below would be altered. Permitting the government to engage in production would involve the addition of a third production sector, a complication which we wish to avoid. A model in which the government purchases capital and labor, using them to produce a flow of government services, is discussed in Sargent and Wallace.

The demand for real money balances is assumed to depend on real GNP, which we take as a proxy for the rate of transacting in the economy, and the after-tax nominal rate of interest, which is the differential between the anticipated return on money holding, $-\pi$, and the anticipated return to bond holding, $(1-t_Y)r-\pi$. Omitting wealth from equation (8) amounts to assuming that households desire to hold any increment in wealth entirely in the form of bonds or equities.[9]

We will usually assume that households' demand for consumption goods is governed by

$$C = c(Y - \delta p_1K - T), \quad 1 > c_Y > 0$$

where

$$Y - \delta p_1K - T = \frac{w}{p}N + (r-\pi)p_1K - T_Y - T_0$$

is householders' real disposable income according to national accounts concepts, although we will also indicate how the model works when more general forms are assumed for the consumption function.[10]

For the economy to be in equilibrium, households' and the government's demands for consumer goods must sum to supply:

$$(10) \quad Y_2 = c(Y - \delta p_1K - T) + g_2$$

Expenditures on consumption goods are

[9] Our assumptions imply that the demand by the public for bonds and equities is given by

$$\left(\frac{B+V}{p}\right)^D = a[Y, (1 - t_Y)r, W],$$

$$a_1 = -m_1, \quad a_2 = -m_2, \quad a_3 = 1$$

where $V = V_1 + V_2$ is the total nominal value of the equities issued by both industries.

[10] One example of such more general forms is

$$C = c\left[Y - \delta p_1K - T - \pi\left(\frac{M+B}{p}\right), (1-t_Y)r - \pi, W\right]$$

$$c_1 > 0, \quad c_2 < 0, \quad c_3 > 0$$

where $\pi(M+B)/p$ is wealth holders' perceived capital loss due to anticipated inflation on government liabilities of fixed nominal value.

assumed to lead to no accumulation of stocks. Instead, consumption goods are assumed to be perfectly perishable.

Other Relationships

Equations (1) through (10) determine the position of the economy at a point in time. In addition to these ten equations, the model contains a few other equations which, while they do not enter into the determination of momentary equilibrium at any particular point in time, are important because they govern the evolution of the system over time. In particular, the economy's rate of capital accumulation is governed by $DK = Y_1 - \delta K - g_1$.

We are assuming that the money wage is rigid at any moment. This assumption is compatible with the existence of a "Phillips curve," for example,

$$\frac{Dw}{w} = \gamma(N, \pi), \qquad \gamma_N > 0, \qquad \gamma_\pi > 0,$$

provided that the implied value of Dw/w is always finite.

Next, we need an equation that explains how expectations of inflation are revised with the passage of time. The "adaptive-expectations" scheme

$$D\pi = \alpha\left(\frac{Dp}{p} - \pi\right), \qquad \infty > \alpha > 0$$

where Dp is a left-hand derivative, is compatible with our assumption that at any moment in time π is fixed.

Finally, we have assumed that firms finance investment with equities. This implies that firms issue equities at a rate with nominal value $DV = p_1 p DK$.

Relationships like these four occupy important roles in determining the behavior of the model over time. However, since this paper is dedicated to the "static" or point-in-time analysis of our model, these equations will be of little concern in what follows.

II. Behavior of the Model

We find it convenient to rewrite the system formed by equations (1) through (10) in the following way:

$$Y_2 = c(Y - T - p_1\delta K) + g_2 \tag{11}$$

$$Y = p_1 Y_1 + Y_2 \tag{12}$$

$$Y_1 = K_1/\beta_1 \tag{13}$$

$$\beta_1 = k_1/f_1(k_1) \tag{14}$$

$$Y_2 = K_2/\beta_2 \tag{15}$$

$$\beta_2 = k_2/f_2(k_2) \tag{16}$$

$$K = K_1 + K_2 \tag{17}$$

$$\frac{M}{p} = m[Y, (1 - t_Y)r] \tag{18}$$

$$f_1'(k_1) = \epsilon \tag{19}$$

$$f_1(k_1) - f_1'(k_1)k_1 = \frac{1}{p_1}\frac{w}{p} \tag{20}$$

$$f_2'(k_2) = p_1\epsilon \tag{21}$$

$$f_2(k_2) - f_2'(k_2)k_2 = \frac{w}{p} \tag{22}$$

$$\epsilon = \delta + \frac{r - \pi}{(1 - t_K)} \tag{23}$$

This is a system of thirteen equations in the thirteen variables Y_1, Y_2, Y, K_1, K_2, k_1, k_2, p, p_1, r, β_1, β_2, and ϵ. The parameters of the model are g_2, T, t_Y, t_K, w, M, π, and K. Our analysis will be static, no time being permitted to pass as we move from one position to another.[11] The variables of the model are assumed to adjust instantaneously in response to the imposition of a disturbance in the form of a change in one or more of the model's parameters.

Our strategy for solving the model is straightforward. We reduce equations (11) through (23) to a pair of relationships

[11] Notice that on our assumptions neither the government's flow budget constraint nor the equation describing the economy's rate of capital accumulation enters as a determinant of the positions of our variables at a particular point in time.

involving Y, real gross national product measured in terms of the consumption good, and ϵ, the gross return to capital in terms of capital goods, and the parameters g_2, T, t_Y, t_K, w, M, π, and K. The first relationship, called the MM schedule, gives the pairs of Y and ϵ which assure portfolio balance for wealth holders. To obtain this relationship we use equations (18), (23), and the profit maximization subsystem, equations (19) through (22). The pairs of ϵ and Y on the MM schedule assure portfolio balance for wealth holders and are compatible with profit maximization, but there is no guarantee that an ϵ and Y pair along the MM curve will result in equality between the stock of capital in existence and the stock firms would like to hold.

The second relationship gives the pairs of ϵ and Y which assure that the market for existing capital is cleared. To obtain this relationship, called the KK schedule, we use all the equations except (18). Pairs of ϵ and Y along this schedule are compatible with profit maximization and assure that firms are content to hold just the existing capital stock; however, wealth holders may not be satisfied with their portfolios at particular Y and ϵ pairs on the KK schedule. The $\epsilon - Y$ pair at the intersection of the KK and MM curves assures both portfolio balance and equilibrium in the market for existing capital.

Some Preliminary Relationships

The first step in obtaining the MM and KK schedules is to use the profit maximization subsystem, equations (19) through (22), to obtain expressions for k_1, k_2, and p_1 in terms of ϵ and an expression for p in terms of ϵ and w. Since $f_1''(k_1)<0$ we can invert (19) to obtain:

$$(24)\quad k_1 = k_1(\epsilon) \quad \text{where} \quad k_{1\epsilon} = \frac{1}{f_1''} < 0$$

We know from profit maximization that the ratio of the marginal product of labor to the marginal product of capital in both industries must equal the wage-rental ratio, which we denote by ω. Thus dividing (20) by (19) and (22) by (21) we have the following:

$$\frac{f_1(k_1)}{f_1'(k_1)} - k_1 = \omega$$

$$\frac{f_2(k_2)}{f_2'(k_2)} - k_2 = \omega$$

$$\omega = \frac{w}{p_1 p\epsilon}$$

Setting the ratios of marginal products in the two industries equal and using (24) we obtain:

$$\frac{f_1[k_1(\epsilon)]}{f_1'[k_1(\epsilon)]} - k_1(\epsilon) = \frac{f_2(k_2)}{f_2'(k_2)} - k_2$$

This equation defines a relationship between k_2 and ϵ. Differentiating and rearranging,[12] we arrive at the following:

$$(25)\qquad k_2 = k_2(\epsilon),$$

$$k_{2\epsilon} = \frac{p_1^2}{f_2''}\,\frac{f_1}{f_2} < 0$$

The economics behind equation (25) is straightforward. Raising the gross rate of return on capital causes the capital-labor ratio in the capital goods industry, k_1, to fall. This decline in k_1 means that the wage-rental ratio, ω, must fall. With a decrease in ω, the ratio of the marginal product of labor to the marginal product of capital in the consumption goods in-

[12]
$$\frac{f_1 f_1''}{(f_1')^2} k_{1\epsilon} d\epsilon = -\frac{f_2 f_2''}{(f_2')^2} dk_2$$

$$dk_2 = \frac{1}{f_2''}\,\frac{f_1}{f_2}\,\frac{(f_2')^2}{(f_1')^2} d\epsilon$$

$$dk_2 = \frac{p_1^2}{f_2''}\,\frac{f_1}{f_2} d\epsilon$$

Thus we can write (25).

dustry must fall in order to reestablish equilibrium, and in order for the ratio of marginal products to fall, the capital-labor ratio in the consumption goods sector, k_2, must decline.

Perfect competition and constant returns to scale imply that each firm in each industry and hence each industry as a whole must earn zero economic profits. We write this condition in intensive form:

$$pp_1f_1 - w - pp_1\epsilon k_1 = 0$$

$$pf_2 - w - pp_1\epsilon k_2 = 0$$

Differentiating these two equations, employing (19) and (21), and rearranging we obtain a relationship between p_1 and ϵ:[13]

$$(26) \qquad p_1 = p_1(\epsilon)$$

where

$$p_{1\epsilon} = \frac{p_1^2}{f_2}(k_1 - k_2) \gtrless 0 \quad \text{as} \quad k_1 \gtrless k_2$$

The effect of an increase in ϵ on p_1 depends upon the relative capital intensities of the two industries. If, as is often assumed, the consumption good is relatively more capital intensive than the capital good, the price of the capital good in terms of the consumption good falls as the real return on capital rises. More generally, we have seen that raising ϵ lowers the wage-rental ratio so that in order for both industries to continue just to cover costs, the industry which uses relatively more capital must have the price of its good rise relative to the price of the good which requires relatively less capital.

Finally, substituting (25) into (22) we have:

$$f_2[k_2(\epsilon)] - f_2'[k_2(\epsilon)]k_2(\epsilon) = \frac{w}{p}$$

Differentiating and rearranging, we obtain:[14]

$$(27) \qquad p = p(\epsilon, w)$$

where

$$p_\epsilon = \frac{p^2p_1^2}{w}\frac{f_1k_2}{f_2} > 0, \quad p_w = \frac{p}{w} > 0$$

The economics embodied in equation (27) is as follows. We know that raising ϵ lowers the capital-labor ratio in the consumption goods industry. This decline lowers the marginal product of labor in that industry, so the real wage in terms of consumption goods must fall. With w fixed, the real wage can fall only if p rises. If w is increased and ϵ stays constant, p must rise in order to keep the real wage the same.

Two other relationships are useful at a later stage. Substituting (24) into (14),

[13] $$(p_1f_1 - p_1\epsilon k_1)dp + (pf_1 - p\epsilon k_1)dp_1 + [(pp_1f_1' - pp_1\epsilon)k_{1\epsilon} - pp_1k_1]d\epsilon - dw = 0$$
$$(f_2 - p_1\epsilon k_2)dp - p\epsilon k_2dp_1 + [(pf_2' - pp_1\epsilon)k_{2\epsilon} - pp_1k_2]d\epsilon - dw = 0$$

Using (19) and (21) and the zero profit condition these two equations become

$$\frac{w}{p}dp + \frac{w}{p_1}dp_1 - pp_1k_1d\epsilon - dw = 0$$
$$\frac{w}{p}dp - \left(\frac{pf_2 - w}{p_1}\right) - pp_1k_2d\epsilon - dw = 0$$

Setting the left-hand sides of these two equations equal to one another we obtain:

$$\frac{w}{p}dp + \frac{w}{p_1}dp_1 - pp_1k_1d\epsilon - dw = \frac{w}{p}dp - \frac{pf_2 - w}{p_1}dp_1 - pp_1k_2d\epsilon - dw$$

$$dp_1 = \frac{p_1^2}{f_2}(k_1 - k_2)d\epsilon$$

Thus we can write (26).

[14] $$-f_2''k_{2\epsilon}k_2d\epsilon = \frac{1}{p}dw - \frac{w}{p^2}dp$$
$$dp = \frac{p^2p_1^2}{w}\frac{f_1k_2}{f_2}d\epsilon + \frac{p}{w}dw$$

Thus we can write (27).

we have a relationship between the capital-output ratio and ϵ:

$$\beta_1 = \frac{k_1(\epsilon)}{f_1[k_1(\epsilon)]}$$

Differentiating and rearranging, we obtain:[15]

$$(28) \qquad \beta_1 = \beta_1(\epsilon)$$

where

$$\beta_{1\epsilon} = \frac{w}{pp_1}\,\frac{1}{f_1^2}\,k_{1\epsilon} < 0$$

Similarly, (16) and (25) yield:

$$(29) \qquad \beta_2 = \beta_2(\epsilon) \text{ where } \beta_{2\epsilon} = \frac{1}{f_2^2}\,\frac{w}{p}\,k_{2\epsilon} < 0$$

In both industries, raising ϵ lowers the capital-output ratio.

The MM Schedule

We are now prepared to derive the *MM* schedule, which shows the pairs of ϵ and Y which are compatible with portfolio balance given that the conditions for profit maximization are satisfied. We use (23) to obtain an expression for the nominal rate of interest in terms of the real gross rate of return ϵ,

$$(30) \qquad r = (1 - t_K)(\epsilon - \delta) + \pi$$

This equation states that the nominal interest rate must be equal to the after-tax real rate of return net of depreciation plus the anticipated rate of inflation. Substituting (27) and (30) into (18), we obtain the *MM* schedule,

$$(31) \qquad \psi(Y, \epsilon; M, t_Y, t_K, \pi) = m[Y, (1 - t_Y)\{(1 - t_K)(\epsilon - \delta) + \pi\}] - \frac{M}{p(\epsilon, w)} = 0$$

We view the *MM* schedule as giving ϵ as a function of Y and the relevant parameters, an appropriate view as long as $\psi_\epsilon \neq 0$. Differentiating (31) and rearranging, we arrive at an expression which allows us to determine both the slope of the *MM* schedule and the shifts in the schedule that are caused by changes in the parameters of the model:

$$(32) \qquad d\epsilon = -\frac{m_1}{\psi_\epsilon}\,dY + \frac{\frac{1}{p}}{\psi_\epsilon}\,dM - \frac{\left(\frac{M}{p^2}\right)p_w}{\psi_\epsilon}\,dw + \frac{m_2 r}{\psi_\epsilon}\,dt_Y - \frac{m_2(1 - t_Y)}{\psi_\epsilon}\,d\pi + \frac{m_2(1 - t_Y)(\epsilon - \delta)}{\psi_\epsilon}\,dt_K$$

$$\psi_\epsilon = \frac{M}{p^2}\,p_\epsilon + m_2(1 - t_Y)(1 - t_K)$$

The signs of the slope of the *MM* schedule and of the expressions for the effects of shifts in the parameters depend crucially on ψ_ϵ. Given the way we have written (31), ψ_ϵ gives the effect of an increase in ϵ on the excess demand for real balances, *ceteris paribus*. Raising ϵ has two offsetting effects on the excess demand for real balances. On the one hand raising ϵ raises r, which makes wealth holders desire smaller real balances. This effect, which we shall call the interest elasticity of demand effect (*IEDE*), tends to reduce excess demand for real balances and is represented by $m_2(1-t_Y)\,(1-t_K)$. On the other hand, raising ϵ means that p must

[15] $$d\beta_1 = \frac{f_1 - f_1' k_1}{f_1^2}\,k_{1\epsilon}d\epsilon$$

$$d\beta_1 = \frac{w}{pp_1}\,\frac{1}{f_1^2}\,k_{1\epsilon}d\epsilon$$

Thus we can write (28).

rise if the profit maximization conditions are to continue to be satisfied. This increase in p reduces the supply of real balances. This effect, which we shall call the profit maximizing supply effect (*PMSE*), and which is represented by $(M/p^2)p_\epsilon$, tends to create an excess demand for real balances. If the *IEDE* is stronger ($\psi_\epsilon<0$), then an increase in ϵ leads to a net decline in the excess demand for real balances and if the *PMSE* is stronger ($\psi_\epsilon>0$), to a net increase in excess demand. We call the case when the *IEDE* is stronger the interest elastic case (*IEC*) and the case when the *PMSE* is larger the interest inelastic case (*IIC*). We believe that the *IEC* is the more important of the two cases, so we will explore only this case in detail. However, we will supply the reader with all the ingredients for a full analysis of the interest inelastic case.

The slope of the *MM* curve and the effect on the position of the *MM* curve of changes in the parameters can now be easily determined. As Y is increased, wealth-holders wish to hold more real balances, creating an excess demand for real balances. Whether ϵ must rise or fall depends upon whether raising ϵ creates an excess supply or demand for real balances when Y is constant. In the *IEC*, ϵ must rise in order to reestablish portfolio balance. This case is shown in Figure 1.

From (31) we have:

$$\left.\frac{d\epsilon}{dY}\right|_{MM}^{IEC} = -\frac{m_1}{\psi_\epsilon} > 0$$

Of course, in the interest inelastic case the *MM* schedule slopes downward.

Our *MM* schedule somewhat resembles the textbook *LM* schedule.[16] The primary difference is that we have insisted that the profit maximization conditions be fulfilled at every point on the *MM* curve. This requirement is not imposed when deriving the usual version of the *LM* curve. It is our new requirement which has raised the ambiguity about the slope of the *MM* curve.

[16] For example, see Martin Bailey.

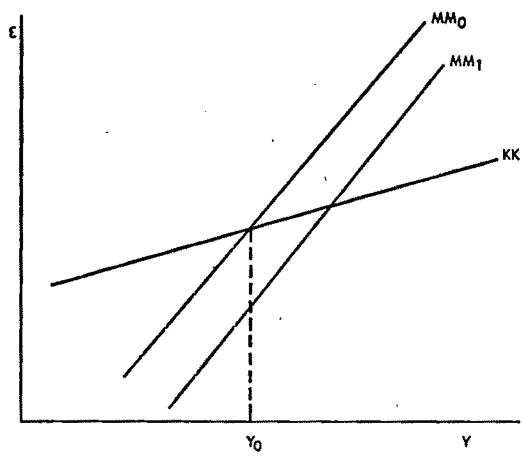


FIGURE 1

We now consider the effects of shifts in the parameters on the *MM* curve.

An increase in the money supply

With Y fixed an increase in M leads to an excess supply of real balances, so in the *IEC* we must have a fall in ϵ in order to reduce the nominal rate so that portfolio balance can be reestablished:

$$\left.\frac{d\epsilon}{dM}\right|_{MM,dY=0}^{IEC} = \frac{\frac{1}{p}}{\psi_\epsilon} < 0$$

Thus an increase in the money supply shifts *MM* down, say from MM_0 to MM_1 in Figure 1.

An increase in the money wage

An increase in the money wage rate, w, means that p must be increased for profit maximization. An increase in p leads to an excess demand for real balances at the original ϵ, so with Y fixed in the *IEC*, ϵ

must rise in order to raise the nominal rate so that portfolio balance can be reestablished:

$$\left.\frac{d\epsilon}{dw}\right|_{MM,dY=0}^{IEC} = -\frac{\frac{M}{p^2}p_w}{\psi_\epsilon} > 0$$

An increase in w shifts the MM curve upward, say from MM_1 to MM_0 in Figure 1.

An increase in t_Y

If there is an increase in t_Y with total tax revenues being held constant by an alteration in T_0, the attractiveness of the return producing assets, bonds and equities, is reduced. At the original ϵ, excess demand for money results, so with Y constant we must have an increase in ϵ in the interest elastic case:

$$\left.\frac{d\epsilon}{dt_Y}\right|_{MM,dY=0}^{IEC} = \frac{m_2 r}{\psi_\epsilon} > 0$$

The MM schedule thus shifts upward with a rise in t_Y.

An increase in anticipated inflation

A rise in the anticipated rate of inflation, π, means that equities become more attractive relative to money, which earns no nominal return, and bonds, the returns from which are fixed in money terms. With ϵ fixed, the nominal rate r rises by $d\pi$, but with Y fixed there will be an excess supply of real balances at this new nominal rate. In the interest elastic case, there must be a decline in ϵ large enough to lower the nominal rate enough to remove the excess supply:

$$\left.\frac{d\epsilon}{d\pi}\right|_{MM,dY=0}^{IEC} = -\frac{m_2(1-t_Y)}{\psi_\epsilon} < 0$$

An increase in π shifts the MM curve down. Given Y, the nominal rate ends up lower after an increase in π, as can be seen from the following expression derived by differentiating (30):

$$\left.\frac{dr}{d\pi}\right|_{MM,dY=0}^{IEC} = \left[1+(1-t_K)\left.\frac{d\epsilon}{d\pi}\right|_{MM,dY=0}^{IEC}\right]$$

$$\left.\frac{dr}{d\pi}\right|_{MM,dY=0}^{IEC} = \frac{\frac{M}{p^2}p_\epsilon}{\psi_\epsilon} < 0$$

The nominal rate ends up lower because as ϵ falls to reestablish equilibrium, p also falls, adding somewhat to the excess supply of money which must be eliminated.

An increase in t_K

If t_K rises, at the original ϵ, the return on equities falls, and since bonds are perfect substitutes for equities, the nominal interest rate falls. At the new nominal rate there will be an excess demand for money, and again assuming that Y is fixed, in the interest elastic case there must be an increase in ϵ in order to raise the nominal rate by enough to remove the excess demand:

$$\left.\frac{d\epsilon}{dt_K}\right|_{MM,dY=0}^{IEC} = \frac{m_2(1-t_Y)(\epsilon-\delta)}{\psi_\epsilon} > 0$$

An increase in t_k shifts the MM curve up. The nominal rate ends up higher after an increase in t_k, as can be seen from the following expressions derived by differentiating (30):

$$\left.\frac{dr}{dt_K}\right|_{MM,dY=0}^{IEC} = -(\epsilon-\delta)+(1-t_K)\frac{d\epsilon}{dt_K}$$

$$= -\frac{(\epsilon-\delta)\frac{M}{p^2}p_\epsilon}{\psi_\epsilon} > 0$$

The KK Schedule

We now proceed to derive the KK schedule, the schedule which shows the

pairs of ϵ and Y that clear the market for the existing capital stock given that the conditions for profit maximization are satisfied. First we obtain expressions for the capital required in each of the industries. From (13) and (15) we have $K_1=Y_1\beta_1$ and $K_2=Y_2\beta_2$. Using (12) we have $Y_1=(Y-Y_2)/p_1$. Thus (17), which states that the market for existing capital is cleared, can be written as follows:

$$K=\left(\frac{Y-Y_2}{p_1}\right)\beta_1+Y_2\beta_2$$

Multiplying through by p_1 we obtain the condition that the total requirements of capital measured in terms of consumption goods must equal the stock of capital goods measured in terms of consumption goods:

$$p_1K=\beta_1(Y-Y_2)+p_1\beta_2Y_2$$

Using (11) and recalling that p_1, β_1, and β_2 can be written as functions of ϵ, we obtain the equation which defines the KK schedule:[17]

$$\begin{aligned}(33)\quad &\Phi(Y,\epsilon;K,g_2,T)\\ &\quad=p_1(\epsilon)K-\beta_1(\epsilon)Y\\ &\qquad-[p_1(\epsilon)\beta_2(\epsilon)-\beta_1(\epsilon)]\\ &\qquad\cdot[c(Y-T-p_1\delta K)+g_2]=0\end{aligned}$$

We view the KK schedule as giving ϵ as a function of Y and the relevant parameters of the model, which requires that $\Phi_\epsilon\neq0$.

[17] Clearly at this stage the reader is free to substitute his favorite consumption function plus g_2 for Y_2. One alternative consumption function is given in fn. 10. In general we could take Y_2 to be a function of Y and ϵ given K and a vector of other parameters. We designate this vector of other parameters α with elements α_i:

$$Y_2=Y_2(Y,\epsilon;K,\alpha)$$

The more general version of (33) is given by the following:

$$\begin{aligned}\Phi(Y,\epsilon;K,\alpha)&=p_1(\epsilon)K-\beta_1(\epsilon)Y\\ &\quad-[p_1(\epsilon)\beta_2(\epsilon)-\beta_1(\epsilon)][Y_2(Y,\epsilon;K,\alpha)]=0\end{aligned}$$

Differentiating (33), we obtain an expression from which we can determine the slope of the KK schedule and the effects of shifts in the parameters on the position of the KK schedule:[18]

$$\begin{aligned}(34)\quad d\epsilon&=\frac{(1-c_Y)\beta_1+c_Yp_1\beta_2}{\Phi_\epsilon}dY\\ &\quad-\frac{p_1+(p_1\beta_2-\beta_1)c_Yp_1\delta}{\Phi_\epsilon}dK\\ &\quad+\frac{(p_1\beta_2-\beta_1)}{\Phi_\epsilon}dg_2\\ &\quad-\frac{(p_1\beta_2-\beta_1)c_Y}{\Phi_\epsilon}dT\\ \Phi_\epsilon&=-[p_1Y_1\beta_{1\epsilon}+p_1Y_2\beta_{2\epsilon}]\\ &\quad+[K_1+(p_1\beta_2-\beta_1)c_Y\delta K]p_{1\epsilon}\end{aligned}$$

The signs of the slope of the KK schedule and of the effects of shifts in the parameters on the position of the KK schedule depend on the sign of Φ_ϵ. Because of the way we have written (33), Φ_ϵ represents the effect of a rise in ϵ on excess supply in the market for the existing stock of capital goods measured in consumption goods, *ceteris paribus*. A rise in ϵ lowers the capital-output ratios in both industries thus leading to an increase in the excess supply of capital goods; this effect, which we call the capital using effect (KUE), is given by $-[p_1Y_1\beta_{1\epsilon}+p_1Y_2\beta_{2\epsilon}]$.

The relative price p_1 may either rise or fall with an increase in ϵ depending upon the relative capital intensities of the two industries; p_1 rises ($p_{1\epsilon}>0$) if $k_1>k_2$ and falls if $k_2>k_1$. Ignoring temporarily the effects on depreciation, we can see from (33) that a rise in the relative price of capi-

[18] If we had used a more general expression for Y_2, as in fn. 17, the differential of the KK schedule would be

$$d\epsilon=-\frac{\Phi_Y}{\Phi_\epsilon}dY-\frac{\Phi_K}{\Phi_\epsilon}dK-\sum_{i=1}^{n}\frac{\Phi_{\alpha_i}}{\Phi_\epsilon}d\alpha_i$$

tal goods, p_1, raises excess supply in the market for existing capital goods. Raising p_1 increases the consumption good value of the supply of existing capital by $Kp_{1\epsilon}$. However, with Y and Y_2 fixed, the consumption good value of demand for existing capital rises by only $K_2p_{1\epsilon}$ since p_1Y_1 must remain constant. Thus excess supply in the market for existing capital rises by $K_1p_{1\epsilon}$. We call this effect the capital valuation effect (*KVE*). The argument is reversed if p_1 falls.

Excess supply also changes because of changes in the value of depreciation. This effect is represented by the term $(p_1\beta_2 - \beta_1)c_Y\delta Kp_{1\epsilon}$. The value of depreciation rises if p_1 rises when ϵ is increased. If the value of depreciation rises, then Y_2 falls, decreasing demand for capital; with a fixed Y this means that p_1Y_1 must rise, increasing the demand for capital. Whether or not there is a net increase in demand depends upon which industry has the higher capital-output ratio; that is, upon the sign of $p_1\beta_2-\beta_1$. Excess supply falls when p_1 rises if $p_1\beta_2-\beta_1<0$. It can be shown that $p_1\beta_2-\beta_1 \gtreqless 0$ as $k_2 \gtreqless k_1$.[19] This means that $p_{1\epsilon}$ and $p_1\beta_2-\beta_1$ always have opposite signs, so the effect of the change in the value of depreciation no matter what its direction is always to reduce excess supply in the market for existing capital. We call this effect the depreciation output switching effect (*DOSE*).

[19] Using Euler's theorem and the profit maximizing conditions we have

$$p_1f_1 = \frac{w}{p} + p_{1\epsilon}k_1$$

$$f_2 = \frac{w}{p} + p_{1\epsilon}k_2$$

Dividing the first equation by k_1 and the second by k_2, subtracting the second equation from the first, collecting terms, and multiplying through by $\beta_1\beta_2$ we obtain

$$(p_1\beta_2 - \beta_1) = \frac{w}{p}\frac{1}{f_1f_2}(k_2 - k_1) \gtreqless 0$$

as $k_2 \gtreqless k_1$.

When $k_1>k_2$ so that $p_{1\epsilon}>0$, Φ_ϵ is positive if the *KUE* plus the *KVE* is greater than the absolute value of the *DOSE*. When, as is often assumed, $k_2>k_1$ so that $p_{1\epsilon}<0$, Φ_ϵ is positive if the *KUE* is greater than the sum of the absolute values of the *KVE* and the *DOSE*. No matter whether $k_1>k_2$ or vice versa, Φ_ϵ is more likely to be positive the more alike are the two sectors, for as $k_1 \to k_2$, we have $p_{1\epsilon} \to 0$, $p_1\beta_2 \to \beta_1$, so the *KVE* and the *DOSE* tend to disappear. Since we believe that the case in which Φ_ϵ is positive is the more important case, we concern ourselves primarily with that case. Since the always positive part of Φ_ϵ depends on substitution in production, we call the case when Φ_ϵ is positive the strong substitution case (*SSC*) and the case when Φ_ϵ is negative the weak substitution case (*WSC*).

Given the sign of Φ_ϵ we can readily find the signs of the slope of the *KK* curve and of the effects on the position of the *KK* curve of changes in the parameters. As Y is increased both Y_2 and p_1Y_1 must rise. These increases generate an increase in the demand for capital. In the strong substitution case, ϵ must rise in order to choke off this excess demand and reestablish equilibrium in the market for existing capital goods. This case is shown in Figure 1.

From (34) we have:

$$\left.\frac{d\epsilon}{dY}\right|_{KK}^{SSC} = \frac{(1 - c_Y)\beta_1 + c_Yp_1\beta_2}{\Phi_\epsilon} > 0$$

Of course, in the weak substitution case the *KK* curve slopes downward.

Reference to (34) makes it easy to determine the effects on the *KK* schedule of changes in the parameters K, g_2, and T.

An increase in the capital stock

At a given level of *GNP*, the effect of an increase in K on the ϵ that clears the market for existing capital and guarantees that

profits are maximized is given by

$$\left.\frac{d\epsilon}{dK}\right|^{SSC}_{KK,dY=0} = -\frac{[p_1+(p_1\beta_2-\beta_1)c_Y p_1\delta]}{\Phi_\epsilon} < 0$$

We are assuming that the numerator of the above derivative is positive, which will be true if δ is sufficiently small or if k_1 is not very much greater than k_2. An increase in K increases the excess supply of capital at a given Y, which means that, in the strong substitution case, ϵ (and therefore p_1p and p) must fall if the market for existing capital is to be cleared. An increase in K thus shifts the KK curve downward.

An increase in government purchases of consumer goods

At a given level of GNP, the effect of an increase in g_2 on the ϵ that clears the market for existing capital and guarantees that profits are maximized is given by

$$\left.\frac{d\epsilon}{dg_2}\right|^{SSC}_{KK,dY=0} = \frac{(p_1\beta_2-\beta_1)}{\Phi_\epsilon} \gtrless 0 \quad \text{as} \quad k_2 \gtrless k_1$$

At a given level of Y, and at the ϵ on the original KK schedule, the output of the consumption-goods sector must expand by dg_2, requiring the output of investment goods to contract by dg_2. At the original levels of p_1, p, and ϵ, firms in the consumption goods sector require $p_1\beta_2dg_2$ additional units of capital to produce the additional output, while capital-producing firms willingly release β_1dg_2 units of capital in order to lower their output by dg_2. If $p_1\beta_2dg_2$ equals β_1dg_2, the market for existing capital remains cleared at the old p_1, p, and ϵ. If $k_2>k_1$, implying that $p_1\beta_2>\beta_1$, the consumption sector demands more additional capital at the original levels of p, ϵ, and p_1p than the capital-producing firms are willing to sell or rent. Thus, there is an excess demand for existing capital, and ϵ (and also p_1p and p) must rise to clear the market for existing capital, given Y. This implies that an increase in g_2 shifts the KK curve upward if $k_2>k_1$. If $k_2<k_1$, the KK curve shifts downward when g_2 increases, while, as we have seen, the KK curve remains unchanged when $k_1=k_2$.

An increase in total real tax collections

The effect of a change in T on the value of ϵ that, given Y, lies on the KK schedule is given by

$$\left.\frac{d\epsilon}{dT}\right|^{SSC}_{KK,dY=0} = \frac{-(p_1\beta_2-\beta_1)c_Y}{\Phi_\epsilon} \gtrless 0 \quad \text{as} \quad k_1 \gtrless k_2$$

An increase of dT has the same effect on the KK schedule as a decrease in government expenditures of (dT/c_Y). The increase in T causes the demand for consumption goods to fall, thereby causing the capital-market-clearing value of ϵ to fall, stay the same, or rise depending on whether k_2 exceeds, equals, or falls short of k_1.

Full Equilibrium

We are now in a position to describe the momentary equilibrium of our economy and the manner in which it responds to the imposition of disturbances. Equilibrium values of ϵ and Y are determined at the intersection of the MM and KK curves, as in Figure 1. Where both the KK and the MM curves are upward sloping, as we have been assuming for the most part, the MM curve must be steeper than the KK curve if the equilibrium is to be stable, a condition that we will assume to be satisfied.[20]

We now describe the effects on momentary equilibrium of changes in the parameters M, w, π, t_Y, t_K, g_2, and T. These effects are summarized in Table 1.

[20] Stability also requires that, in a certain sense, price adjustments in the market for paper assets are swift relative to price adjustments in the market for existing physical capital. A formal discussion of stability is contained in the Appendix.

TABLE 1—SUMMARY OF RESULTS[a]

Effect on	Increase in M	w	t_Y	t_K	π	g_2: $k_2>k_1$ IEC	g_2: $k_2>k_1$ IIC	g_2: $k_1>k_2$ IEC	g_2: $k_1>k_2$ IIC
Y	+	−	−	−	+	+	−	−	+
N	+	−	−	−	+	?	?	?	?
p	+	−	−	−	+	+	+	−	−
pp_1	+	−	−	−	+	+	+	−	−
ϵ	+	−	−	−	+	+	+	−	−
r	+	−	−	−	+	+	+	−	−

[a] The results reported in this table are based on the assumption that the KK schedule has a positive slope and on the assumption in fn. 22 about the effect of a rise in ϵ on employment. The results reported in columns headed by IEC are based on the assumption that the MM schedule has a positive slope. The results reported in columns headed by IIC are based on the assumption that the MM schedule has a negative slope. The results for an increase in T with t_Y and t_K constant are just the reverse of those for an increase in g_2.

An increase in the money supply

An increase in the money supply shifts the MM curve downward, indicating that ϵ must fall if portfolio balance is to be maintained at the original level of GNP, say Y_0 in Figure 1. To equilibrate the money market, r and ϵ tend to fall. Perceiving that wealth-holders are willing to accept a lower rate of return on equities, firms now find it profitable to use more capital intensive techniques and bid for capital goods. This causes the money price of capital goods, p_1p, to rise. Since the money wage rate is unchanged and since firms must continue just to cover costs, p must rise when p_1p rises.[21] The rise in p both reduces the supply of real balances and raises the level of output that profit-maximizing firms want to produce with the existing capital stock. These two effects tend to create excess demand for real balances, which means that r and ϵ must rise. The rise in p also increases the real rate of return to capital implied by profit maximization. Our assumption that the MM schedule is steeper than the KK schedule guarantees that as output and prices rise, the portfolio-balancing ϵ eventually "catches up" with the capital-market-clearing ϵ, establishing a new equilibrium at higher values of both ϵ and Y, as in Figure 1. The rise in ϵ is necessitated by the higher marginal product of capital that prevails at the new equilibrium. The increase in the money supply also produces rises in N, r, p, and pp_1.[22] Whether or not

21
$$\frac{d\{p_1[\epsilon(p,w)]p\}}{dp} = p\frac{dp_1}{dp} + p_1$$

where $\epsilon(p, w)$ is obtained by "inverting" (27). Thus

$$\frac{dp_1(p)p}{dp} = \frac{w}{pk_2f_1}[k_1 - k_2] + p_1$$
$$= \frac{w}{pk_2f_1}k_1 + \frac{k_2(pp_1f_1 - w)}{pk_2f_1} > 0$$

That is, through equation (20) a rise in p_1p causes a fall in k_1 which through equation (19) leads to a rise in ϵ. This in turn requires a fall in k_2 through equation (21), which necessitates a fall in the real wage in terms of the consumption good. Thus p must rise.

22 The effects on employment produced by changes in the parameters can be seen by analyzing the following expression for labor employed, N:

$$N = \lambda_1(\epsilon)\left\{\frac{Y - c[Y - T - p_1(\epsilon)\delta K] - g_2}{p_1(\epsilon)}\right\} + \lambda_2(\epsilon)\{c[Y - T - p_1(\epsilon)\delta K] + g_2\}$$

where $\lambda_1 = 1/f_1$, $\lambda_2 = 1/f_2$ and $\lambda_{1\epsilon}, \lambda_{2\epsilon} > 0$

Upon differentiation we have:

p_1 rises or falls depends on relative capital intensities.

An increase in the money wage

An increase in the money wage acts exactly like a decrease in the money supply. The increase in w causes the MM curve to shift up, since for profit-maximization, p must rise, creating excess demand for money at the original Y and ϵ, tending to drive ϵ upward. The tendency for ϵ to rise causes firms to attempt to rent or sell capital, causing p_1p, p, ϵ, and output to fall. Equilibrium is reestablished at values of ϵ and Y lower than those characterizing the original equilibrium.

An increase in anticipated inflation

An increase in the expected rate of inflation π shifts the MM curve downward, eventually resulting in increases in both ϵ and Y. At the original level of GNP, the increase in π reduces the ϵ required for portfolio balance. This causes firms to bid for capital, causing p_1p, p, ϵ and Y to increase.

An increase in the income tax rate with total real tax collections held constant

An increase in t_Y reduces the attractiveness of bonds and equities, creating an excess demand for money at the original ϵ and Y. At the original Y, r and ϵ must rise in order to equilibrate the money market; that is, the MM schedule shifts upward. The rise in ϵ causes firms to attempt to rent or sell capital, causing p_1p, p, ϵ, and Y to fall until equilibrium is reestablished at values of ϵ and Y that are lower than those characterizing the original equilibrium.

An increase in the profit tax rate with total real tax collections held constant

An increase in t_K causes the MM schedule to shift upward, thereby depressing ϵ and Y. The rise in t_K means that at the initial Y the original ϵ now implies a lower rate of return to equities, causing wealth-holders to attempt to move into money. This creates a tendency for r and ϵ to rise at the original Y, causing firms to attempt to dispose of capital, driving p_1p, p, ϵ, and Y downward.

An increase in government purchases of consumer goods

The effects of an increase in g_2 depend on the relative capital intensities of the two sectors. If $k_2 > k_1$, the increase in g_2 increases the capital-market-clearing value of ϵ. This occurs because at the original Y

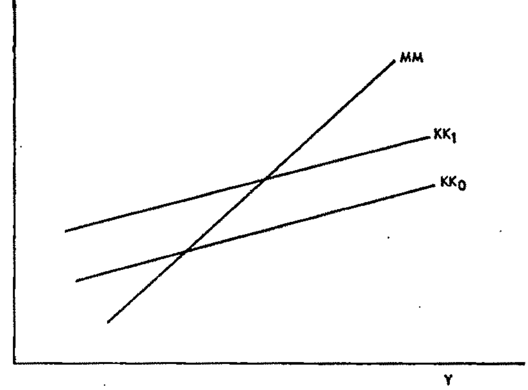


FIGURE 2

$$dN = \left\{ Y_1\lambda_{1\epsilon} + Y_2\lambda_{2\epsilon} - \frac{1}{p_1}\left[N_1 + (p_1\lambda_2 - \lambda_1)c_Y\delta K\right]p_{1\epsilon} \right\} d\epsilon + \frac{1}{p_1}\left[(1-c_Y)\lambda_1 + c_Y p_1\lambda_2\right]dY + \frac{1}{p_1}(p_1\lambda_2 - \lambda_1)dg_2 - \frac{1}{p_1}(p_1\lambda_2 - \lambda_1)c_Y dT - \frac{1}{p_1}(p_1\lambda_2 - \lambda_1)c_Y p_1\delta dK$$

Raising ϵ with Y constant increases the labor-output ratio in both industries which tends to increase employment. An increase in ϵ also gives rise to two effects analogous to the KVE and the $DOSE$ effects pertinent for evaluating Φ_ϵ above. These two effects may tend to decrease employment as ϵ is increased. However, these effects become smaller as $k_1 \to k_2$, and we assume that k_1 and k_2 are close enough together that a rise in ϵ increases employment. A rise in output with ϵ constant always increases employment. Since ϵ and Y always move together when changes occur in M, w, t_Y, t_K, and π, N always moves in an unambiguous direction given our assumptions.

and ϵ the rise in g_2 increases the amount of capital demanded by the consumption goods sector by more than the capital producing firm can release and still have Y remain constant. This causes firms to bid for capital, causing p_1p, p, ϵ, and Y to increase. On the other hand, if $k_1>k_2$, an increase in g_2 shifts the KK schedule downward, causing Y and ϵ to fall. In both cases the effect on employment is ambiguous.[23] If $k_1=k_2$, an increase in g_2 has no effects on Y, N, p, ϵ, or p_1. Of course, N_1, N_2, K_1, and K_2 will change according to

$$dK_2 = \beta_2 dg_2$$

$$dK_1 = -\frac{\beta_1}{p_1}dg_2 = -dK_2$$

$$dN_2 = \frac{\beta_2}{k_2}dg_2$$

$$dN_1 = -\frac{\beta_1}{k_1p_1}dg_2 = -dN_2$$

In this case, market forces induce producers of capital goods to release just the amounts of capital and labor that producers of consumer goods need to meet the increased demand for their output. At the initial equilibrium values of Y, Y_1, and Y_2, the increased demand for consumer goods caused by the increase in g_2 produces a tendency for p to rise, while p_1p remains constant. The rise in p induces a rise in profits of consumer goods producers, while profits of investment goods producers remain unchanged:

$$\left.\frac{d\sum_1}{dp}\right|_{dp_1p=0} = 0$$

$$\left.\frac{d\sum_2}{dp}\right|_{dp_1p=0} = (1-t_K)Y_2 > 0$$

The rate of return on capital in the consumer goods industry rises, exceeding that in the investment goods industry. This causes firms in the consumer goods industry to bid for capital, which firms in the investment goods industry are willing to sell or rent to them in order to obtain the highest possible rate of return on their capital. The slightest upward pressure on p is thus sufficient to bring about any required changes in the distribution of capital and employment. Consequently in equilibrium, p and p_1 remain at their initial levels, as do Y and N.[24]

An increase in total real tax collections, tax rates being held constant

An increase in T with t_y and t_K held constant operates in the same fashion as a

[23] The expression in fn. 22 reveals the source of this ambiguity. When g_2 is increased, ϵ and Y move in the same direction if the MM curve slopes upward. When $k_2>k_1$, ϵ and Y both rise. However, an increase in g_2 also has an employment-reducing effect in addition to the employment-increasing effects of the increases in ϵ and Y. At a constant Y, a rise in g_2 shifts production from the good with the higher labor-output ratio, Y_1, to the good with the lower labor-output ratio, Y_2, thus reducing employment. A similar argument reveals an ambiguity in the effect of a rise in g_2 on employment when $k_1>k_2$.

[24] Notice that we are assured that, at the initial equilibrium values of p_1, p, r, and Y, the government is able to finance the increase in the deficit caused by the increase in g_2. Substituting the national income identity and the government budget constraint into the definition of disposable income, and noting that saving plus consumption equals disposable income, we arrive at

$$S = p_1DK + \frac{DM}{p} + \frac{DB}{p}$$

where S is saving and p_1DK equals the real value of the rate at which firms are issuing equities to finance their capital accumulation. The above equation states that in equilibrium the rate at which households are actually acquiring equities, bonds, and money must equal the rate at which they desire to save. Where capital intensities are equal, the increase in g_2 produces an equal reduction in p_1DK and hence in the rate at which firms issue equities. This is precisely enough of a decrease to offset the increase in the rate at which the government must issue bonds and/or money, since $dg_2=-p\,dDK$, and since from the government budget constraint and the fact that prices don't jump,

$$dg_2 = d\left(\frac{DM}{p}+\frac{DB}{p}\right)$$

decrease in g_2. With an upward sloping MM curve, an increase in T drives Y and ϵ up (down) when $k_1>k_2(k_2>k_1)$.

The "Interest Inelastic Case"

The analysis above assumes that the interest elasticity of the demand for money is sufficiently high in absolute value to make the MM curve upward sloping. We now briefly consider the interest inelastic case, in which the MM curve slopes downward. We continue to assume that the KK schedule slopes upward.

Reference to equation (32), which gives the differential of the MM schedule, establishes that when the MM schedule is downward sloping, the MM schedule is shifted *upward* by *increases* in the money supply and the anticipated rate of inflation, and by *decreases* in w, t_K, and t_y. It follows that in the interest inelastic case, changes in M, w, π, t_K, and t_y exert effects on Y, ϵ, p, and p_1p that are qualitatively the same as those in the interest elastic case in which the MM schedule is upward sloping.

Where the possibility of a downward sloping MM curve does create some ambiguity is in the analysis of changes in g_2 and T. Suppose that $k_2>k_1$ and that the government increases its purchase of consumer goods, causing KK to shift upward, as in Figure 3. At the original Y, the capital-market-clearing value of ϵ rises, causing firms to bid for capital, raising p_1p and p. The rise in p diminishes the supply of real balances tending to create excess demand for real balances. Since the interest elasticity of the demand for money is relatively small, large increases in r, and therefore in the ϵ required by wealthholders, are required in order to diminish the demand for real balances by a given amount. The ϵ consistent with portfolio balance tends to rise above the value of ϵ that with the new KK schedule clears the capital market at the old level of output. This causes firms to attempt to dispose of

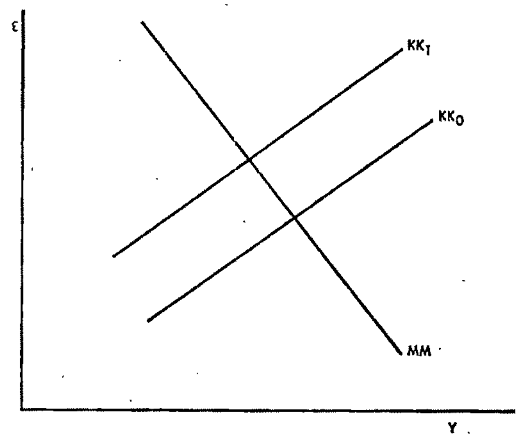


FIGURE 3

capital exerting downward pressure on p_1p, p, ϵ, and output. Output and ϵ fall until eventually equilibrium is restored at a lower Y but at a higher ϵ than characterized the initial equilibrium.

The upshot is that in the interest inelastic case, an increase in g_2 decreases the level of output if $k_2>k_1$. Similarly, an increase in g_2 increases (leaves unchanged) output if $k_1>k_2$ $(k_1=k_2)$. Once again the effect on employment is ambiguous no matter which good is more capital intensive.[25]

III. Conclusions

Increases in the money supply and the expected rate of inflation and decreases in the money wage and the marginal income and profits tax rates increase output, employment, and prices in our model. Those changes have their effects by introducing disturbances in the markets for stocks of physical and paper assets, thereby exerting

[25] The expression in fn. 22 reveals the source of this ambiguity When g_2 is increased, ϵ and Y move in opposite directions if the MM curve slopes downward. When $k_2>k_1$, ϵ rises, and Y falls. The rise in ϵ tends to increase employment; the fall in Y tends to reduce it. As mentioned in fn. 23 there is an additional employment reducing effect caused by a shift in output between sectors. A similar argument reveals an ambiguity in the effect of a rise in g_2 on employment when $k_1>k_2$.

pressure on prices, causing firms to adjust their levels of output and employment. On the other hand, increases in the level of government expenditures and decreases in tax collections may either increase or decrease output and employment depending upon the relative capital intensities of the two production sectors and some other conditions involving the interest elasticity of the demand for money. These fiscal policies affect output and employment only to the extent that consumption goods and capital goods are not perfect substitutes in production; if they are perfect substitutes, the rates of government expenditures and of tax collections play no role in the determination of output and employment at a point in time.

These properties of our model serve to highlight just how important are the characteristics of the market in existing physical capital. The assumption made here is that a "perfect" market exists, one in which firms can buy and sell capital with zero transactions costs. This implies that profit-maximizing firms will hold the optimal quantity of capital at each instant. Firms' trading in the market for existing capital will thus guarantee that the profit-maximizing marginal conditions for capital are met at each instant, just as the marginal conditions for labor are assumed to be met in the standard Keynesian model. This makes the market for existing capital a critical element in the mechanism by which policy changes and other disturbances affect the economy.

On the other hand, the Keynesian model denies the presence of a market in existing capital and permits discrepancies to exist at any moment between the marginal product of capital and the cost of capital. High transactions costs are assumed to prevent firms from adjusting their capital stocks instantaneously. Instead, they are assumed to respond to an increase in the gap between the marginal product of capital and the cost of capital, as might be caused by a decline in the interest rate, by increasing the rate at which they are investing. Thus, the Keynesian model posits investment demand curves for firms rather than insisting that the marginal conditions for capital always be satisfied, as we have done here. That feature of the Keynesian model is an extremely important one in making variations in government expenditures a potent instrument for stabilization with predictable effects on output and employment in the short run.

Appendix

Equations (11) to (23) can also be reduced to a system of two equations in p and ϵ. The implicit relationship between ϵ, p, and w used to derive (27) can also be used to write the following:

$$(A1) \quad \hat{\epsilon} = \hat{\epsilon}(p; w) \quad \text{where} \quad \hat{\epsilon}_p = \frac{1}{p_\epsilon} > 0$$

$$\hat{\epsilon}_w = p_w \frac{wf_2}{p^2 p_1^2 f_1} < 0$$

$\hat{\epsilon}$ is the rate of return on capital in terms of capital goods implied by profit maximization at a given p. We now use ϵ to denote the rate of return on capital which is necessary for portfolio balance. Equilibrium requires that the rate of return on capital implied by profit maximization must equal the rate of return on capital required for portfolio balance:

$$(A2) \quad \Gamma(\epsilon, p; w) = \hat{\epsilon}(p; w) - \epsilon = 0$$

Using (A1) we can rewrite (33) in the following way:

$$\begin{aligned} \theta(Y, p; w, K, g_2, T) &= p_1(\hat{\epsilon})K - \beta_1(\hat{\epsilon})Y \\ &\quad - [p_1(\hat{\epsilon})\beta_2(\hat{\epsilon}) - \beta_1(\hat{\epsilon})] \\ &\quad \cdot \{c[Y - T - p_1(\hat{\epsilon})\delta K] + g_2\} = 0 \end{aligned}$$

Since $\theta_Y \neq 0$, we can use this implicit relationship between Y and p to solve for Y in terms of p and the relevant parameters:

$$(A3) \quad Y = Y(p; w, K, g_2, T)$$

where

$$Y_p = -\frac{\theta_p}{\theta_Y} = -\frac{\Phi_\epsilon \frac{1}{p_\epsilon}}{\Phi_Y} = \left\{\frac{1}{\left.\frac{d\epsilon}{dY}\right|_{KK}}\right\}\frac{1}{p_\epsilon}$$

Y_p has the same sign as the slope of the KK curve, so, given the assumptions in the text, Y_p is positive. Substituting (A3) into (18) and using (23) we arrive at the following:

$$\text{(A4)}\quad \Omega(\epsilon, p; w, K, g_2, T, M, t_Y, t_K, \pi) = m\{Y(p; w, K, g_2, T), (1-t_Y)\cdot[(1-t_K)(\epsilon-\delta)+\pi]\} - \frac{M}{p} = 0$$

The description of the pressures toward adjustment given in the text can be written down explicitly in the following pair of differential equations (A5) and (A6).

Employing the usual technique, we linearize the system around the equilibrium values for ϵ and p which we call ϵ_0 and p_0. See equation (A7) below. All the derivatives are evaluated at (ϵ_0, p_0). The characteristic equation of the system is

$$\text{(A8)}\quad \lambda^2 + B\lambda + C = 0$$

$$B = \left\{\mu_1\left[\frac{r}{(1-t_Y)(1-t_K)}\right]\cdot\left[\frac{m_2(1-t_K)(1-t_Y)}{M/P}\right] + \mu_2\left[\frac{pp_1}{p_1+pp_{1\epsilon}/p_\epsilon}\right]\left[\frac{1}{\epsilon p_\epsilon}\right]\right\}$$

$$C = \mu_1\mu_2\left[\frac{r}{(1-t_Y)(1-t_K)}\right]\cdot\left[\frac{pp_1}{p_1+pp_{1\epsilon}/p_\epsilon}\right]\left[\frac{p}{M\epsilon p_\epsilon}\right]$$

$$\text{(A5)}\quad \frac{Dr(\epsilon)}{r(\epsilon)} = \mu_1\left[\frac{m\{Y(p; w, K, g_2, T), (1-t_Y)[(1-t_K)(\epsilon-\delta)+\pi]\} - \frac{M}{p}}{\frac{M}{p}}\right]$$

$$\text{(A6)}\quad \frac{D(pp_1[\hat{\epsilon}(p; w)])}{pp_1[\hat{\epsilon}(p; w)]} = \mu_2\left\{\frac{\hat{\epsilon}(p)-\epsilon}{\hat{\epsilon}(p)}\right\}$$

$$\text{(A7)}\quad \begin{bmatrix} \frac{(1-t_Y)(1-t_K)}{r}(D\epsilon - D\epsilon_0) \\ \\ \frac{p_1+pp_{1\epsilon}\frac{1}{p_\epsilon}}{pp_1}(Dp-Dp_0) \end{bmatrix} = \begin{bmatrix} \mu_1\frac{m_2(1-t_K)(1-t_Y)}{\frac{M}{p}} & \mu_1\left\{\frac{\frac{M}{p^2}+m_1Y_p}{\frac{M}{p}}\right\} \\ \\ \mu_2\left(-\frac{1}{\epsilon}\right) & \mu_2\left\{\frac{\frac{1}{p_\epsilon}}{\epsilon}\right\} \end{bmatrix}\begin{bmatrix} \epsilon-\epsilon_0 \\ \\ p-p_0 \end{bmatrix}$$

$$\cdot\left[m_2(1-t_K)(1-t_Y)+\frac{M}{p^2} + m_1\frac{1}{\left.\frac{d\epsilon}{dY}\right|_{KK}}\right]$$

Necessary and sufficient conditions for stability are $B>0$ and $C>0$. Our assumption that markets adjust instantaneously implies that $\mu_1, \mu_2 \to \infty$. For stability we must have μ_1 and μ_2 approach infinity in such a way that B remains positive. The following condition must always hold:

$$\frac{\mu_2}{\mu_1} < \frac{\frac{p}{M} r \,|\, m_2 \,|}{\left[\frac{pp_1}{\epsilon(p_1 p_\epsilon + p p_{1\epsilon})}\right]}$$

The ratio of the speed of adjustment in the market for existing capital to the speed of adjustment in the money market must be less than the ratio of the effect of a rise in ϵ on the percent excess demand for money to the effect of a rise in p on the percent gap between the profit-maximizing and the portfolio-balancing rate of return on capital. Adjustment must be relatively more rapid in the markets for paper assets. This condition is necessary because a rise in p raises $\hat{\epsilon}$, thus increasing the gap between the profit-maximizing and portfolio-balancing rates of return. If μ_2 is small enough, where "small enough" is defined by the condition above, to keep this disequilibrating effect from becoming dominant, then the system as a whole is stable.

For C to be positive the following condition must hold:

$$m_2(1-t_K)(1-t_Y)+\frac{M}{p^2}\,p_\epsilon+m_1\left(\frac{1}{\left.\frac{d\epsilon}{dK}\right|_{KK}}\right)>0$$

This condition can be rewritten as the following:

$$-\frac{1}{\left.\frac{d\epsilon}{dY}\right|_{MM}}+\frac{1}{\left.\frac{d\epsilon}{dK}\right|_{KK}}>0$$

With our assumptions that

$$\left.\frac{d\epsilon}{dK}\right|_{KK}>0$$

stability is assured if

$$\left.\frac{d\epsilon}{dY}\right|_{MM}<0$$

If

$$\left.\frac{d\epsilon}{dY}\right|_{MM}>0$$

then we must have

$$\left.\frac{d\epsilon}{dY}\right|_{MM}>\left.\frac{d\epsilon}{dY}\right|_{KK}$$

for stability.

REFERENCES

M. J. Bailey, *National Income and the Price Level: A Study in Macroeconomic Theory*, 2d edition, New York 1971.

D. K. Foley and M. Sidrauski, *Monetary and Fiscal Policy in a Growing Economy*, New York 1971.

D. W. Jorgenson, "The Theory of Investment Behavior," in R. Ferber, ed., *Determinants of Investment Behavior*, New York 1967.

J. M. Keynes, *The General Theory of Employment, Interest, and Money*, London, New York 1936.

F. Modigliani and M. H. Miller, "The Cost of Capital, Corporation Finance, and the Theory of Investment," *Amer. Econ. Rev.*, June 1958, *48*, 261–97.

T. Sargent and N. Wallace, "Market Transaction Costs, Asset Demand Functions, and the Relative Potency of Monetary and Fiscal Policy," *J. Money, Credit, and Banking*, suppl. May 1971, *3*, 469–505.

J. Tobin, "A Dynamic Aggregative Model," *J. Polit. Econ.*, Apr. 1955, *63*, 103–115.

H. Uzawa, "On a Two-Sector Model of Economic Growth," *Rev. Econ. Stud.*, Oct. 1961, *28*, 40–47.

N. Wallace, "A Static Nonstationary Analysis of the Interaction Between Monetary and Fiscal Policy," Univ. Minnesota Center for Economic Research disc. paper no. 9, 1971.

The Demand for State and Local Government Employees

By RONALD G. EHRENBERG*

During the 1953–70 period, state and local government employment grew faster than total employment, rising from 8.6 percent to 14.0 percent of total employment on payrolls of nonagricultural establishments. While much of the rise was due to the growth of educational employment (145 percent), state and local noneducational employment also grew significantly faster (79 percent) than employment in the private nonagricultural economy (33 percent). Indeed, several noneducational employment categories have grown at rates approaching or exceeding that of educational employment in recent years (Table 1). By 1980, state and local government employment is projected by the Bureau of Labor Statistics to increase further to 15.9 percent of total nonagricultural payroll employment.[1]

Interest in the determinants of the level of such employment within a state and its allocation across functional categories is more than academic. Recently, fears have been expressed that the growing strength of public employee unions, their increased militancy, and the trend toward collective bargaining rights with respect to wage issues, will lead to inflationary wage increases in the sector and aggravate the financial problems facing state and local governments.[2] These fears are explicitly based upon the assumption that public services are essential and consequently that the demand for public employees is wage inelastic. To many, the logical conclusion then drawn is that limitations should be placed on the collective bargaining rights of these groups.[3]

To date, however, no estimates have been presented of the *wage elasticities* of demand for public employees and consequently no quantitative discussion of the strength of the market forces, which might prevent such increases, has occurred. To the extent that state and local government employee unions value both the wage and employment levels of their members, high wage elasticities coupled with projected relatively slower rates of growth of employment demand would help to moderate the size of wage increases which they will seek in the future.

One can easily think of possibilities for substituting capital for labor in the provision of public services (for example, police patrol cars) and for substituting private for public provision (for example, private security agencies). Moreover, given the limited resources which state and local

* Associate professor of economics, University of Massachusetts. Without implicating them for what remains, I am grateful to Orley Ashenfelter, James Kindahl, a referee, and the editor for their comments on earlier versions of the paper. The research was supported by a faculty research grant from the Manpower Administration, U.S. Department of Labor and the material presented here draws from my larger study (1972), prepared for that organization.

[1] See *Manpower Report*, Tables C-1, E-11; *Handbook of Labor Statistics*, Tables 38, 100; *Public Employment* (various issues).

[2] Harry Cohany and Lucretia Dewey, Sheila White' and Sheila Weissbror present descriptions of these trends. Table 1 presents data which indicate that both over the 1961–70 and 1965–70 periods, state and local government employees' wages have increased more rapidly than wages in the private sector. However, as suggested by the editor, the past relative inflation of public employees' wages was probably due to the rapid growth of employment demand in the sector and unrelated to the wage elasticities of demand.

[3] See, for example, Harry Wellington and Ralph Winter.

TABLE 1—TOTAL PERCENTAGE CHANGES FULL-TIME EQUIVALENT STATE AND LOCAL GOVERNMENT EMPLOYMENT AND MONTHLY PAYROLL PER MAN, BY FUNCTION

	Employment		Payroll Per Man	
Category	1965–70	1961–70	1965–70	1961–70
Education	27.6	60.5	41.8	66.5
Highway	3.4	10.6	38.5	61.6
Public Welfare	47.1	99.5	41.6	68.2
Hospitals	18.6	38.3	51.4	74.4
Health	36.0	55.3	43.6	68.6
Police	29.0	44.9	43.0	75.8
Fire Protection	12.1	18.4	47.7	74.9
Sewerage	10.9	26.4	37.4	59.0
Other Sanitation	11.4	17.9	37.7	59.6
Parks and Recreation	17.2	29.8	34.4	56.7
Natural Resources	12.6	23.6	41.2	65.3
Correction	28.0	54.6	44.9	76.6
Libraries	9.9	N.A.	42.2	N.A.
Employment Security Administration	18.0	N.A.	33.6	N.A.
Financial Administration	12.6	19.0	41.1	61.3
General Control	31.6	45.4	38.2	62.8
Local Utilities	7.0	14.8	38.0	76.2
Private-Nonfarm	14.2	27.9	31.4	50.5
Private-Manufacturing	7.4	18.8	28.7	44.8

Source: Derived from data found in various issues of *Public Employment* and *Handbook of Labor Statistics*. The private figures are percentage changes in total payroll employment and average hourly earnings of nonsupervisory employees.

governments can command, an increase in the relative price of a public service should lead to a substitution against that service. Thus, it is not obvious, a priori, that the wage elasticities of demand for all categories of state and local government employees are inelastic.

The primary purpose of this paper is to present empirical estimates of the wage elasticities of demand for different categories of state and local government employees. The employment demand equations that are estimated are derived from a utility maximization model of state and local government behavior. After presenting this model in the first section, we next briefly discuss the data used in the study. The structural system of demand equations is then estimated using pooled time-series and cross-section information, with annual individual state data as the units of observation. A number of alternative estimation methods are used in the analysis. Parameter estimates obtained from the model are utilized in the final section to simulate the *disemployment effects* of postulated future relative wage increases for individual classes of state and local government employees, as well as increases for all classes relative to the private sector.

Our estimates indicate that statistically significant negative wage elasticities of demand exist for almost all functional categories of state and local government employees. However, these elasticities do appear to be inelastic and consequently the simulated disemployment effects are quite small. Thus, while there is evidence that state and local government decision makers respond to market forces in choosing their employment portfolios, these responses do not appear to be sufficiently

strong to limit the size of real wage increases which state and local government employees may seek in the future.

I. Theoretical Structure

Abstracting from the myriad of state and local government bodies within a state, we consider a single decision-making unit which is assumed to determine the allocation of state resources between public and private production and also the per capita level and functional distribution of state and local government employment. Employing a "utility-tree" approach, let the utility function of this unit be written as

$$(1)\quad U = U\left[V^A\begin{bmatrix}\text{privately}\\ \text{produced}\\ \text{goods and}\\ \text{services}\end{bmatrix}, V^B\begin{bmatrix}\text{publicly}\\ \text{produced}\\ \text{goods and}\\ \text{services}\end{bmatrix}\right]$$

The decision-making body is assumed to derive utility from the various categories of public services, with the per capita service flow from each category being proportional to the per capita category employment level. At each point in time, for any function, the factor of proportion is assumed constant across states. Hence, we are implicitly assuming no quality differentials within an employment category across states and, since state price indices for the "cost of capital" do not exist, we have also ignored the possibility of substituting capital for labor.

Specifically, suppose that the branch utility function (V^B) is of the form,[4]

$$(2)\quad V^B = V^B\left[\frac{M_1}{P} - \frac{b_1}{P}, \frac{M_2}{P} - \frac{b_2}{P}, \ldots, \frac{M_n}{P} - \frac{b_n}{P}\right]$$

$$M_j > b_j \text{ for all } j$$

Here, M_k is the full-time equivalent state and local government employment level of the kth functional category, b_k is the minimum required number of type k employees, P is the state population, and there are n functional employment categories.

The minimum required employment levels are introduced because state and local government decision making is often alleged to be conducted on an incremental basis, with previous departmental budgets or employment levels taken as given.[5] To allow for this possibility we assume, in a manner analogous to that of Pollack and Wales, that the minimum required number of employees in a category is proportional to the lagged category employment level.

$$(3)\quad b_k = \alpha_k M_k^{t-1}, \quad 0 \leq \alpha_k \leq 1, \quad k = 1, 2, \ldots, n$$

Employment decisions are strictly incremental if the α_k are all unity, while the entire employment portfolio is a decision variable if all the α_k are zero. Intermediate values indicate that the category employment levels can be cut to some extent. An interpretation of (2) and (3) is then as follows: For any level of budgeted funds, the decision-making body first makes expenditures for the minimum required numbers of employees and then with the remaining budgeted funds, chooses the increments to these employment levels so as to maximize

[4] A specific member of this class of utility functions is the Stone-Geary utility function,

$$U(M) = \sum_{k=1}^{n} a_k \log\left(\frac{M_k}{P} - \frac{b_k}{P}\right), \quad a_k > 0, \quad \sum_{k=1}^{n} a_k = 1$$

Richard Stone, Richard Parks, and Robert Pollack and Terrance Wales utilize this utility function to derive and estimate consumer demand functions.

[5] Ira Sharkansky and the references cited therein present evidence that agencies are seldom cut below their previous levels of appropriations. However this in itself is not evidence of incremental budgeting, for with growing state and local government budgets it is conceivable that the appropriate marginal conditions would dictate that all agencies be expanded anyway.

the branch utility function (V^B) which depends only on the per capita increments.

Suppose that the decision makers seek to maximize the "higher-level" utility function (1) subject to a constraint that involves total available resources in the state, here assumed to be a function of per capita personal income and per capita grants from the federal government, and the prices of the various individual publicly and privately produced services. If we further assume that the higher-level utility function is strongly separable, which requires that the marginal rates of substitution between different categories of public services be independent of "consumption" levels of privately produced services, then Robert Strotz has shown that the decision makers will optimally follow a two-stage process.[6] First the expenditure levels for each aggregated branch (public vs. private) will be selected with the resources allocated to each branch being a function only of the total resources available and the aggregate branch price indices. Then expenditure levels for specific commodities, in either the public or private branch, will be determined as functions of the total own branch expenditure level and the specific prices of all commodities within the branch alone.

If we assume that the per capita total state and local government employment budget in a state is approximately proportional to the per capita "public branch allocation," the above suggests that it is reasonable to specify the per capita total employment budget (B/P) as being determined by,[7]

$$\log(B/P) = S_0 + S_1 \log(RW) + S_2 \log(G/P) + S_3 \log(Y/P) \tag{4}$$

Here Y/P is per capita personal income, G/P is per capita grants from the federal government to state and local governments in the state and RW is a measure of the average relative cost of publicly produced to privately produced goods and services. For empirical purposes RW is defined as the ratio of average monthly earnings of state and local government employees to the comparable figure for manufacturing production workers in the state.

Once the per capita total employment budget is determined, the decision makers choose the per capita level and functional distribution of state and local government employment by maximizing the branch utility function (2) subject to (3) and the constraint that the total employment budget be exhausted.

$$\sum_{k=1}^{n} W_k(M_k - b_k) = B - \sum_{k=1}^{n} W_k b_k \tag{5}$$

W_k is the monthly payroll cost per employee of the kth category. Assuming that the customary regularity conditions hold for (2), its maximization subject to (3) and (5) yields a general system of category employment demand equations.

$$\frac{M_j^t}{P} - \frac{\alpha_j M_j^{t-1}}{P} = F\left(W_1, W_2, \ldots, W_n, \left(B - \sum_{k=1}^{n} W_k \alpha_k M_k^{t-1}\right) \Big/ P\right) \quad j = 1, 2, \ldots, n \tag{6}$$

In our empirical work (6) is approxi-

[6] As we have already indicated in our introduction, the strong separability assumption is not strictly valid as many services produced in the public sector have viable private substitutes. Furthermore, privately produced but publicly purchased services would be included in branch A. Nevertheless, the assumption is useful in that it provides an analytic foundation for (4).

[7] Decision makers are viewed in (4) as having sufficient flexibility so that even grants tied to specific functions serve to increase the total employment budget, rather than only the specific category budget. Moreover, matching grants are assumed to have only income and not price effects. A long literature exists (for example, Edward Gramlich and David Smyth) which neither conclusively support nor refute these assumptions. Federal grants are also treated as being exogenous for our purposes. Our formulation is dictated by data considerations and the particular functional form is chosen for analytic convenience.

mated by the following system of log-linear employment demand equations.[8]

$$(7)\quad log\left(\frac{M_j^t}{P}-\frac{\alpha_j M_j^{t-1}}{P}\right)=b_{0j}+b_{1j}\,log\,W_j$$
$$+b_{2j}\,log\left(B-\sum_{k=1}^{n}W_k\alpha_k M_k^{t-1}\right)\Big/ P$$
$$+\sum_{R=1}^{m}c_{Rj}\,log\,SDR \qquad j=1,2,\ldots,n$$

To avoid problems of severe multicollinearity, (7) excludes all wage variables except the employment category's own wage. Furthermore, since it is erroneous to specify that the relative importance of the various services are constant across states or within a state over time, the system includes as explanatory variables those sociodemographic variables (*SDR*) that are expected to influence the decision makers' tastes for the various services.[9] On the basis of data availability and preliminary experimentation, the variables utilized are the state population density (*SD*1), the proportion of the state population ages 5–17 (*SD*2), and the proportion of the state population over age 65 (*SD*3). The system of $n+1$ equations contained in (4) and (7) is estimated in Section III.[10]

[8] The specific utility function found in fn. 4 leads directly to the system of demand equations in (7), with the further restrictions that $b_{1j}=-1$ and $b_{2j}=1$ for all j. Since we are primarily interested in ascertaining the magnitude of the wage and employment budget elasticities of demand, these restrictions are not imposed in the estimation and the specific functional form of the utility function is not assumed a priori. However, we emphasize that (7) must then be regarded only as an approximation to a true system of demand equations.

Note that if government employers face positively sloped labor supply curves, the wage elasticities estimated from (7) will be biased in a positive direction. Orley Ashenfelter presents evidence, however, that the aggregate supply of labor to the state and local sector is highly elastic and thus this bias is likely to be small.

[9] For example, we would expect the age distribution of the population to effect the relative importance to the decision makers of educational employment. Similarly population density will influence the relative importance of police employment.

[10] I also estimated a variant of (7) which substituted the category wage relative to the average wage for all state and local government employees in the state (W_j/W_T) for the category own wage (W_j). The estimates were similar in nature to those obtained using specification (7) and are omitted here for brevity. See Ehrenberg, chs. 4 and 5.

II. The Data

Published data allow us to consider eleven functional categories of employees and to pool time-series and cross-section information, with annual individual state data during the 1958–69 period as the units of observation.[11] Full-time equivalent state and local government employment has been reported by function and state since 1953, however the number of functional categories reported declined gradually from nineteen in 1969 to ten in 1953. Thus increasing the length of the time period under consideration can be done only at the expense of aggregating categories.

The choice of 1958 as the initial year was made on pragmatic grounds; by the availability of data on per capita federal grants to state and local governments. Referring to Table 1, this required that sewerage and other sanitation employment be aggregated into a sanitation category and financial administration employment be included with general control. Due to the similarity of functions performed by the two categories in each of these aggregate groups there appeared to be obvious reporting errors over time within several states (for example, the ratio of financial administration to general control employment in a state fluctuated widely over time, while the total of the two and their relative wage remained fairly constant). Hence both of these aggregations appeared to be desirable on a priori grounds. Finally, I constructed a miscellaneous all others category that included parks and recreation, correction, library, employment se-

[11] The variables used in the study are constructed from published data found in the U.S. government publications which are listed in the references.

curity administration, local utility, state liquor store, and all other state and local government employees not explicitly contained in any of the categories listed in Table 1.[12]

All of the variables expressible in dollar terms were deflated by the national consumer price index to obtain real values; unfortunately no individual state price indices exist and we were unable to control for interstate price differences. The average monthly payroll cost per man variable was derived by dividing full-time equivalent employment into total monthly payroll. In 1964 the reported full-time equivalent employment variable was rounded to the nearest 100 employees, which induced large errors in the wage variable for small states and/or relatively small functional employment categories. Consequently, data from 1964 was eliminated from the sample. Finally, no data was reported for Alaska or Hawaii in 1958 and 1959. The final sample thus consists of 546 state-year observations.

III. Empirical Estimates

This section presents estimates of the demand for state and local government employees. We first briefly discuss estimates of the derived reduced form model, when the minimum required employment level for each class is assumed to be zero, and consider the legitimacy of aggregating the data across years. Maintaining the same assumption, we next present estimates of the structural model utilizing the pooled data. Finally, the incremental budgeting model is estimated under a restrictive set of assumptions.

[12] For a description of the type of functions performed by employees in each category and an attempt to use a clustering procedure to determine the appropriate aggregated groups, see Ehrenberg, ch. 3. A more complete description of the methods of, and problems involved in constructing the variables utilized in the study is also found there.

Estimates of the Reduced Form and Tests for Aggregation Bias

Initially we assume that the minimum required employment level for each class of employees is zero; that the decision maker is free to choose the entire employment portfolio each period. Substituting (4) into (7) yields the reduced form system of demand equations. This system was estimated in three different ways by ordinary least squares: First, separately for each year in the sample using cross-section data. Second, using the pooled data for all years. Third, using the pooled data with a set of zero-one dummy variables included to allow the intercept term to vary over time.[13]

In general, the estimated coefficients of the economic variables in the individual year cross-sections, especially the wage elasticities, were statistically insignificant. In those cases that a particular coefficient was significant, it tended to vary widely across years. This clearly points out the dangers inherent in drawing conclusions from a single year's cross-section study.

As expected the estimates based upon the pooled data were vastly superior in terms of the signs and statistical significance of the key economic variables.[14] However one may question the legitimacy of aggregating the data across years. Using the sum of squared residuals obtained from the individual year and pooled regressions, F-statistics were calculated to test for the possibility of aggregation bias.[15] Specifi-

[13] The intercept dummy variables are included to capture intertemporal productivity variations, since if the productivity of state and local government employees varies over time, the demand for each category will also vary, *ceteris paribus*. Inasmuch as their inclusion did not significantly alter the estimated wage elasticities of demand, the intercept shift variables are not utilized in the subsequent analysis.

[14] For brevity, these estimates are not presented here since an analysis of them would be quite similar to that of the structural estimates.

[15] See Frank Fisher.

cally we consider two null hypotheses here:

(1) That the entire set of reduced form coefficients is identical across years for each category.

(2) That the entire set of reduced form coefficients except the intercept is identical across years for each category.

For six of the eleven functional categories the null hypothesis that the entire set of reduced form coefficients is equal across years was rejected. Furthermore, the rejection was not due solely to the intercept term varying; in four of these six cases null hypothesis (2) was also rejected.

For pragmatic reasons once again, we will continue to utilize the pooled data in what follows. However, our results must be qualified by noting that for about half of the categories information will be lost by the aggregation. Our simulations in the next section will be based in these cases upon estimated *average* elasticities which may differ from the marginal elasticities which we would obtain if significant coefficients could be derived from the individual year cross-section regressions.

Structural Estimates—No Incremental Budgeting

To estimate the structural model, we first estimate (4) to obtain an estimated per capita total employment budget ($\hat{B}/P$) and then use this as an instrumental variable in the estimation of (7). Moreover, if the employment budget is fixed, a positive residual in spending for one functional category must be offset by a negative residual in some other category. Due to the contemporaneous correlation of residuals across equations in (7), one would expect three-stage least squares estimates ($3SLS$) to be asymptotically more efficient than two-stage least squares estimates and the former method is used below.[16]

Utilizing the pooled data, estimation of (4) yielded the following (where the numbers in parentheses are absolute t-values).

$$\text{(8)}\quad \begin{aligned} \log(B/P) = &\underset{(21.48)}{-4.120} + \underset{(7.87)}{.315}\log(RW) \\ &+ \underset{(17.32)}{.222}\log(G/P) \\ &+ \underset{(28.60)}{.753}\log(Y/P) \\ &\qquad R^2 = .754 \end{aligned}$$

The less than unitary elasticity for RW indicates that, *ceteris paribus*, an increase in the average wage of public employees relative to private sector wages will cause a disemployment impact as the per capita total employment budget would increase by a smaller percentage. As indicated in Section IV, however, this impact may be partially offset if the resulting adjustments in the labor market lead to an increase in per capita income. Furthermore, to simulate the estimated impact on each category's employment level, requires estimation of the complete system.

The $3SLS$ estimates of (7) obtained under the assumption that the minimum required employment levels are all zero, are presented in Table 2.[17] The own wage elasticities are negative and statistically significant in all cases, save for the miscellaneous all other category.[18] Except for the public welfare category these elasticities are all significantly less than unity. Thus, as conjectured by many, the *partial* wage elasticities of demand for public employees do appear to be primarily inelastic.[19] Note however, that even the "essential services" police and fire protection have significant,

[16] Two-stage least squares estimates of (7) differed only marginally.

[17] The standard errors of these estimates are asymptotic estimates, hence the t-tests are only asymptotically valid and should be interpreted heuristically.

[18] Given the heterogeneous nature of the functions performed by employees in the latter category, the insignificant elasticity is not unexpected.

[19] They are partial elasticities because they do not consider the impact on the employment budget equation of an increase in a category's wage.

TABLE 2—3*SLS* ESTIMATES—POOLED DATA[a]
LOG-LINEAR EQUATIONS

Dependent variable (M_j/P)	W_j	$(\hat{B}/P)$	$SD1$	$SD2$	$SD3$
Education	−.425 (8.89)[b]	.730 (16.61)	−.055 (10.98)	.183 (3.95)	
Highway	−.594 (11.55)	.441 (8.04)	−.160 (27.48)		
Public Welfare	−1.001 (10.58)	1.569 (19.25)	.029 (2.93)		.289 (6.72)
Hospital	−.410 (9.86)	.571 (7.87)	.058 (5.64)		.353 (8.67)
Health	−.279 (4.08)	.684 (9.06)			−.361 (9.88)
Police	−.281 (5.73)	.795 (16.08)	.064 (11.35)		
Fire	−.277 (5.65)	.850 (12.95)	.133 (15.39)		
Sanitation	−.504 (7.55)	.593 (6.45)	.141 (12.25)		
Natural Resources	−.530 (5.76)	.282 (3.19)	−.334 (36.21)		
General Control	−.316 (7.49)	.820 (19.53)	−.072 (14.23)		
All Other	.067 (1.18)	.740 (12.68)	.008 (1.16)		

Notes: M_j/P = per capita full-time equivalent state and local government employees of category j.
G/P = per capita federal grants to state and local governments, in real terms.
Y/P = per capita personal income in real terms.
B/P = per capita total employment budget for state and local government employees, in real terms.
W_j = average monthly payroll cost per full-time employee in category j, in real terms.
RW = ratio of average monthly payroll cost per man of all full-time state and local government employees to average monthly earnings of manufacturing production workers.
$SD1$ = population density
$SD2$ = proportion of the population ages 5–17.
$SD3$ = proportion of the population, age 65 and up.

Sources: Derived from data in U.S. Bureau of Census, *Estimates of the Population of States by Age*, CRP Series P-25 (various issues); *Government Finances* (various issues); *Public Employment* (various issues); *Statistical Abstract of the United States, 1969; Handbook of Labor Statistics 1970; Survey of Current Business* (Apr. 1965, Apr. 1970).

[a] The number of state-year observations for all equations is 546; 11 years (1958–63, 65–69) times 50 states minus the observations for Alaska and Hawaii in 1958 and 1959.

[b] Absolute t-values are shown in parentheses.

although small, wage elasticities of demand. The estimates of the per capita employment budget elasticities are all positive and statistically significant.

With respect to the sociodemographic variables, population density appears to be negatively related to education, highway, natural resource, and general control employment. The latter relationship may reflect the substitution of capital for labor that is possible in large scale clerical and administrative operations. This variable is also positively related to public welfare, hospital, police, fire, and sanitation employment. Several plausible factors relating to the relationship between population density and the costs and benefits from these services may explain the signs of these coefficients.

As expected, the proportion of the population between the ages of 5 and 17 is positively related to educational employment and the proportion of the population older than age 65 is positively related to public welfare and hospital employment. The negative relationship between the proportion of the population older than age 65 and health employment may be due to the preventive nature of many of the services in this category (for example, school immunization programs); services which are not directed primarily at the aged.[20]

Structural Estimates—Incremental Budgeting

To identify the α_k coefficients that appear in (7) requires either a method which allows non-linear constraints to be placed on parameters across equations or an iterative procedure. Software limitations prevented the former and to iterate over eleven coefficients searching for maximum-likelihood sets of estimates would have proven prohibitively expensive. Consequently, we adopt a second best procedure and assume that the minimum required proportion of last period's per capita employment level (α_k) is the same for all functions (α). While this assumption might be justified by arguing that all agencies apply equal pressure on decision makers to preserve their sizes, its main justification is in its analytic convenience.

Because the system (7) is log-linear, one cannot shift the $(\alpha M_j^{t-1}/P)$ variable in each equation to the right-hand side and thus the dependent variable will vary as α varies. Hence, we were unable to follow a procedure similar to that used by other investigators, of choosing the value of α which minimizes the sum of squared residuals in (7) across all eleven equations,

[20] Two extensions of the analysis that were conducted warrant brief reporting: First, regardless of whether state and local government decision making is actually governed by an explicit utility maximization framework, it seems desirable to use an econometric approach in which the sum of the estimated payroll expenditures across categories adds up to the estimated total employment budget. This can be expressed as the Engel aggregation condition, that the sum of the per capita employment budget (B/P) elasticities of demand, weighted by their respective category proportions of total payroll expenditures, be unity. Formal F-tests of this restriction for the estimates of (7) reported in Table 2, utilizing 1969 national average category expenditure weights, indicated that the weighted sum was significantly less than unity. Consequently, the system (7) was reestimated using a restricted 3*SLS* technique (see Henri Theil), with the Engel aggregation condition imposed as a constraint. The restricted wage elasticities of demand differed only marginally from the unrestricted estimates, except the elasticity for educational employment which doubled in magnitude. The employment budget elasticity for education also increased substantially; intuitively this was expected to occur as the expenditure weight on education is over 50 percent. The employment budget elasticities for the other categories increased only marginally.

Second, the dangers inherent in inferring time-series responses from pooled cross-section time-series data are well known. While the lack of degrees of freedom generally led to statistically insignificant coefficients when the system (4), (7) was estimated *separately* for each individual state using the short time-series 1958–63, 65–69; these statistically insignificant individual state time-series coefficients were used to perform nonparametric tests to ascertain if the wage elasticities of demand for each category appeared to be "primarily" negative as postulated. Such tests (Ehrenberg, ch. 4) supported this hypothesis for most of the categories. See Alexander Mood and Franklin Graybill, p. 403–25 for the basis of the nonparametric test used.

TABLE 3—3SLS ESTIMATES—INCREMENTAL BUDGETING MODEL[a]
LOG-LINEAR EQUATIONS
$\alpha = .5$

Dependent variable $(M_j^t/P) - .5(M_j^{t-1}/P)$	W_j	$(\hat{B}/P)^b$	$SD1$	$SD2$	$SD3$
Education	−.175 (3.15)[c]	.037 (6.01)	−.085 (11.09)	.170 (2.43)	
Highway	−.635 (14.23)	.394 (8.02)	−.176 (25.54)		
Public Welfare	−.338 (3.65)	.888 (10.49)	−.008 (.64)		.452 (7.98)
Hospital	−.516 (7.42)	.223 (2.41)	.038 (2.80)		.379 (6.74)
Health	−.300 (4.27)	.608 (7.56)			−.405 (8.29)
Police	−.023 (.45)	.451 (8.23)	.028 (3.77)		
Fire	−.240 (3.37)	.638 (8.11)	.098 (8.59)		
Sanitation	−.539 (7.54)	.239 (2.49)	.131 (9.56)		
Natural Resources	−.602 (6.37)	.278 (3.12)	−.343 (28.57)		
General Control	−.099 (1.64)	.480 (8.32)	−.101 (12.36)		
All Others	.261 (4.39)	.320 (4.68)	−.023 (2.41)		

[a] For number of observations, sources of data, and definitions, see Table 2.
[b] In this table $(\hat{B}/P)$ is used as a short-hand notation for

$$(\hat{B}/P) - .5 \sum_{j=1}^{11} W_j(M_j^{t-1}/P)$$

[c] Absolute t-values in parentheses.

time periods and states.[21] Rather, 3SLS estimates of the system were derived for a set of values of α and it was heuristically determined that the minimum required proportion (α) lies in the range 0.0 to 0.5.[22] For values of α which fall in this range, *disemployment effects* of postulated future relative wage increases are simulated in the next section. Since these estimates are tolerably stable over the range of values, it is unnecessary for us to choose a "statistically optimal" estimate for α.

For comparison purposes, estimates of (7) when α was assumed to equal .5, are presented in Table 3.[23] In general as α increased from 0 to .5, the employment budget elasticities decreased; however, there was no uniform pattern of change for the own wage coefficients.

[21] See, for example, Stone.

[22] The system (7) was estimated as α varied from 0.0 to 0.9 in steps of 0.1. Two criteria were used in our

IV. Implications For Public Policy

The impact on per capita full-time equivalent state and local government employment in each functional category of an increase in the category real wage can be simulated using the parameter estimates described in the previous section. We consider two cases here:[24]

A. The real wage of a functional category of state and local government employees increases relative to all other employees' real wages and nonlabor real personal income.

B. The real wages of all state and local government employees increase proportionately, relative to all other employees' real wages and nonlabor real personal income.

For both cases, we calculate the *total* own wage elasticity of demand for each functional employment category, which consists of two components. First, the direct *partial* elasticity, based upon the estimated own wage (W_j) elasticities obtained from the individual category employment demand equations. Second, the indirect effect, which operates through the impacts of an increase in the class wage on the relative cost of publicly to privately produced services (RW) and on per capita personal income (Y/P) in the per capita total employment budget equation. Since the per capita total employment budget (B/P) tends to increase due to these impacts, as long as an employment category has a positive employment budget elasticity, its total wage elasticity will be smaller in absolute value than the partial elasticity.[25] Moreover, *ceteris paribus*, the larger the partial wage elasticity, the smaller the employment budget elasticity, and the smaller the share of the total employment payroll spent on the category, the larger in absolute value will be the total wage elasticity. Hence, the larger will be the disemployment impact that would result from an increase in the class wage.

Differentiating the system (4), (7) with respect to the real wage of each class, one can derive formulae for the total own wage elasticities of demand for the two cases considered.[26] In addition to the parameter estimates obtained from the estimation of (4) and (7) these elasticities are seen to depend upon the fraction of the employment budget initially spent on each category, the proportion of personal income earned by state and local goverment employees, the ratio of per capita total state and local government employment lagged

heuristic evaluation of which values of α yielded plausible estimates. First, as observations were eliminated from the sample when $M_j/P-\alpha M_j^{t-1}/P$ was negative for any category, what proportion of the original number of observations remained in the sample for each value of α? Second, for each value, what proportion of the wage and employment budget elasticities were statistically significant and of the anticipated sign? Details of this analysis are also found in Ehrenberg, ch. 4.

[23] It is easily seen that b_{1j} and b_{2j} in (7) represent the steady-state wage and employment budget elasticities, respectively.

[24] Several additional cases are considered in Ehrenberg, ch. 5.

[25] An increase in state and local government employees' wages would lead to the following set of adjustments in the labor market; a decline in state and local government employment (the impact which we seek to estimate) and an increase in the level of unemployment and/or an increase in the supply of labor to the private sector, resulting in lower wages but greater employment there. To the extent that private sector wages and employment do not change (perhaps due to wages which are inflexible in a downward direction), per capita personal income would increase as long as the demand for state and local government employees is wage inelastic. Our simulations ignore the adjustments that would occur in private sector wages and employment (i.e., they assume a private sector wage elasticity of unity), and may to a minor degree overstate the impact of public employee wage changes on per capita personal income As a consequence, the total elasticities presented in Table 4 may be slightly biased towards zero. However, since the partial elasticities provide an absolute upper bound for the total elasticities, our main conclusions would be unaltered. The simulations also assume that per capita federal grants are unrelated to state and local government employees' wages.

[26] The formulae and their derivations are available upon request from the author.

one year to the current total, and the comparable ratio for per capita employment in each category. National average values for each of these variables in 1969 are used in the simulation.

Substitution of these values and the necessary parameter estimates obtained from the estimation of (4) and (7) into the derived formulae, yields empirical estimates of the total own wage elasticities of demand. These are calculated for seven different sets of estimates of (7).[27] In Table 4, for each functional category and case, we indicate the interval over which the seven sets of estimates varied. It should be emphasized that these do *not* represent statistical confidence intervals, rather they indicate only the sensitivity of the total elasticities to the value of α utilized. The short-run elasticities reported in Table 4 refer to the impact in the first year, while the long-run elasticities indicate the steady-state solutions. The two are identical when $\alpha=0$.

Considering first case A, in which employees in a functional category succeed in pushing up their real wages relative to all other employees and nonlabor personal income, the estimates appear to be tolerably stable for each category, except education and public welfare. Moreover in all cases, save perhaps public welfare, there is no doubt that the total own wage elasticities are *inelastic*. A 5 percent increase in the real wage of a class would, *ceteris paribus*, decrease *per capita* full-time equivalent in the class by less than 5 percent. Furthermore, since population growth has averaged more than 1 percent a year, the category decline in *total* full-time-equivalent employment would be even smaller.

Due to the fact that an increase in any class wage increases the sum of the committed employment expenditures, all partial cross-wage elasticities of demand are negative in the short-run for case A. However, in most cases the magnitudes of these partial cross wage-elasticities are extremely small, often less than $-.001$.[28] Furthermore, when we consider the indirect impacts through the total employment budget equation, the short-run total cross-wage elasticities are seen to be primarily positive and in the long-run they are all positive.

Turning to case B, in which all state and local government employees' real wages increase proportionately relative to all other employees' wages and nonlabor personal income, the estimated disem-

TABLE 4—DISEMPLOYMENT IMPACT SIMULATIONS: INTERVAL ESTIMATES OF TOTAL OWN WAGE ELASTICITIES

Category	Short Run	Long Run
A.		
Education	−.56, −.09	−.57, −.08
Highway	−.59, −.32	−.64, −.44
Public Welfare	−1.13, −.18	−1.13, −.33
Hospital	−.46, −.26	−.51, −.30
Health	−.31, −.16	−.32, −.26
Police	−.31, −.02	−.35, −.01
Fire	−.28, −.12	−.31, −.23
Sanitation	−.57, −.28	−.56, −.40
Natural Resources	−.59, −.29	−.60, −.39
General Control	−.34, −.05	−.34, −.09
All Others	.03, .16	.30, .29
B.		
Education	−.36, .04	−.36, .02
Highway	−.42, −.04	−.48, −.26
Public Welfare	−.46, .46	−.46, −.00
Hospital	−.19, −.06	−.42, −.15
Health	−.02, .28	−.06, −.01
Police	.00, .30	.00, .15
Fire	.03, .34	−.03, .13
Sanitation	−.29, −.07	−.45, −.25
Natural Resources	−.44, −.09	−.49, −.14
General Control	−.02, .28	−.03, .08
All Others	.30, .42	.30, .45

[27] These estimates correspond to values of α assumed equal to 0 (Table 2), 0—restricted estimates, .1, .2, .3, and .4 (all not presented here for brevity), and .5 (Table 3).

[28] For example, when $\alpha=.5$, only 6 of 110 partial cross-wage elasticities of demand were greater than .1 in absolute value with all referring to employment responses to changes in the education wage.

ployment effects are now all smaller. Inasmuch as a proportionate state and local government employee wage increase removes the incentive to substitute across functional categories, this result was exexpected. Those elasticities that remain negative are all inelastic; however, several of the intervals now contain zero and a few of the estimated elasticities are even always positive. For example, in the long run, police, fire, general control, and the miscellaneous all other category, all appear to have positive total wage elasticities of demand.

These results suggest that while state and local governments do respond to market forces in choosing their employment portfolios, these market forces do not appear to be sufficiently strong to limit the size of real wage increases which state and local government employees may seek in the future. They also suggest that employees in each functional category have a stake in the economic well-being of members of all the other categories, as each category can minimize the potential employment losses of their members by seeking to increase the real wages of all categories, rather than merely their own real wage. Consequently, continual careful consideration, on both sides, should be given to the evolving structure of collective bargaining in this sector.[29]

[29] The estimates of (4) and (7) can also be used to simulate the impact of public sector employment programs (such as that contained in the *Emergency Employment Act of 1971*) in which the federal government makes cash grants to state and local governments in an attempt to increase the number of public sector jobs. Ehrenberg, ch. 5, presents such a simulation.

REFERENCES

O. Ashenfelter, "Demand and Supply Functions for State and Local Government Employment: Implications for Public Employment Programs," unpublished paper, Princeton, Mar. 1972.

H. Cohany and L. Dewey, "Union Membership Among Government Employees," *Mon. Lab. Rev.*, July 1970, *93*, 15–20.

R. Ehrenberg, *The Demand for State and Local Government Employees: An Economic Analysis*, Lexington 1972.

F. Fisher, "Tests of Equality Between Sets of Coefficients in Two Linear Regressions: An Expository Note," *Econometrica*, Mar. 1970, *38*, 361–66.

E. Gramlich, "The Effect of Federal Grants on State-Local Expenditures": A Review of the Econometric Literature," *Nat. Tax Assn. Proc.*, 1969, *62*, 569–93.

A. Mood and F. Graybill, *Introduction to the Theory of Statistics*, 2d ed., New York 1963.

R. Parks, "Systems of Demand Equations: An Empirical Comparison of Alternative Functional Forms," *Econometrica*, Oct. 1969, *37*, 629–51.

R. Pollack and T. Wales, "Estimation of the Linear Expenditures System," *Econometrica*, Oct. 1969, *37*, 611–28.

I. Sharkansky, *Spending in the American States*, Chicago 1968.

D. Smyth, "The Response of State and Local Governments to Federal Grants," *Nat. Tax J.*, Sept. 1968, *21*, 349–58.

R. Stone, "Linear Expenditure Systems and Demand Analysis: An Application to the Pattern of British Demand," *Econ. J.*, Sept. 1954, *64*, 511–27.

R. Strotz, "Empirical Implications of the Utility Tree," *Econometrica*, Apr. 1957, *25*, 269–80.

H. Theil, *Economic Forecasts and Policy*, 2d ed., Amsterdam 1961.

S. Weissbror, "Changes in State Labor Laws in 1970," *Mon. Lab. Rev.*, Jan. 1971, *94*, 61–62.

H. Wellington and R. Winter, "The Limits of Collective Bargaining in Public Employment," *Yale Law J.*, June 1969, *69*, 1107–27.

S. White, "Work Stoppages of Government Employees," *Mon. Lab. Rev.*, Dec. 1969, *92*, 29–34.

A. Zellner, "An Efficient Method of Estimat-

ing Seemingly Unrelated Regressions and Tests for Aggregation Bias," *J. Amer. Statist. Assn.*, June 1962, *57*, 348–68.

U.S. Bureau of the Census, *Estimates of the Population of States by Age, Current Population Reports,* Series P-25, various issues.

———, *Governmental Finances,* various issues.

———, *Public Employment,* various issues.

———, *Statistical Abstract of the United States,* Washington 1969.

U.S. Bureau of Labor Statistics, *Handbook of Labor Statistics 1970,* Bull. 1666, Washington 1970.

U.S. Department of Commerce, *Surv. Curr. Bus.*, Apr. 1965 and Apr. 1970.

U.S. Department of Labor, *Manpower Report of the President—1971,* Washington 1971.

Quarterly Models of Wage Determination: Some New Efficient Estimates

By J. C. R. ROWLEY AND D. A. WILTON*

The concept of stable Phillips curves and their economic implications have been the subject of a series of disputes in recent years. Since Phillips introduced his simple disequilibrium model, empirical research in the field of wage determination has indicated a wide spectrum of explanatory variables which expands and contracts almost randomly in published studies. Apart from this difficulty in specification, the intertemporal instability of estimated coefficients for Phillips curve variants is particularly disturbing.[1] In addition, the theoretical bases for many of these variants are unclear and several economists have participated in a neoclassical counter-attack on the existence and stability of Phillips curves from a theoretical viewpoint.[2]

The objectives of this paper are to examine the consequences of certain invalid statistical procedures which are employed in many of these studies and to provide some empirical evidence of their effect. In particular, we focus attention on the implications of the aggregative procedure which provides the basis for the use of a "four-quarter overlapping change" representation of the dependent variable in quarterly studies of wage determination and, also, for the use of moving-averages of explanatory variables. A partial list of studies which make use of these specifications would include the following works: Paul Anderson, Ronald Bodkin et al., George de Menil, Leslie Dicks-Mireaux and Christopher Dow, Otto Eckstein, Michael Evans and Lawrence Klein, John Helliwell et al., Klein and James Ball, Edwin Kuh, Michael Levy, Richard Lipsey and Michael Parkin, George Perry, Edmund Phelps, Gale Pierson, Grant Reuber, Charles Schultze and Joseph Tryon, Norman Simler and Alfred Tella, and Wayne Vroman.

Although these forms of variables are predominant in many studies, the implications of the aggregative procedure for the error terms are usually either completely neglected or represented by simple statements of "possible" problems which are not discussed further.[3] Kuh's assertion that " . . . artificial serial correlation is often introduced through the overlapping process which partially neutralizes the confidence to be placed in the larger t-coefficients" (p. 346) is rare in its explicit reference to the problem of autocorrelation. In the first section of the paper we present a set of assumptions sufficient for

* Economic Council of Canada and Statistics Canada, respectively. We are indebted to the Canadian Department of Labour for research support and to the Institute for Economic Research at Queen's University.

[1] For example, in our attempts to reproduce the estimates of the study by Bodkin et al., we mistakenly added two additional observations to our initial sample. The estimated coefficient of the unemployment term increased from 10.4 to 17.3. For further evidence of the inherent temporal instability of Phillips curve variants, see Levy and Perry, pp. 72–84.

[2] See the studies by Phelps et al.

[3] There are exceptions. For example, see Dicks-Mireaux and Dow, Sparks and Wilton, and Sargan. In addition, Kuh and Eckstein present models with both quarterly and overlapping annual changes in wage-rates which indirectly illustrate the nature of the problem. The use of quarterly changes reduces the problem of autocorrelation which we shall illustrate but introduces other difficulties. Student's t-statistics are much lower for models of quarterly changes than for the models of overlapping annual changes.

the specification of the "four-quarter overlapping-change" model in wage analysis. Serial correlation in the error term is shown to be a necessary consequence of the model. However, if the collective explanatory variables can be calculated, then generalized least squares (*GLS*) estimators can likewise be calculated for the coefficients as well as ordinary least squares (*OLS*) estimators. These two sets of estimators provide the data for our assessment of the effect of the aggregative procedure.

Necessary and sufficient conditions for least squares estimators to be fully efficient despite the presence of autocorrelation, so that both standard errors and Student's t-statistics are unaffected by it, have been established by William Kruskal, Sujit Mitra and C. Radnakrishna Rao, Rao (1967, 1968) and George Zyskind (1962, 1967). These statistical results need to be supplemented in particular analyses by empirical results since they indicate that optimality of least squares procedures may depend upon both the particular sample to be used and the weights adopted for the aggregative procedure. If the least squares estimates of coefficients are not optimal, conventional estimates of standard errors and Student's t-statistics are based on inappropriate formulae. However, the parametric estimates remain unbiased and the inaccuracies of these formulae may be small. Their extent must be established empirically, although Benee Swindel and Geoffrey Watson (1955, 1967) provide theoretical accounts of bounds on these inaccuracies. Three tables, which contain *OLS* and *GLS* results for three different studies of wage determination, reveal that the use of these formulae leads to substantial errors in inference.

As a basis for discussion, the study by Perry provides an appropriate framework since it is one of the earliest and, perhaps, best elucidated studies on the use of overlapping annual changes in the context of quarterly data. We have presented a complete analysis of the general model of discontinuous wage changes elsewhere (1971) so only a brief statement of the set of sufficient conditions for the four-quarter overlapping change model will be presented below. Most of these assumptions are either explicitly or implicitly stated by Perry (pp. 30–31).

I. The Model of Wage Determination

(A1) Wages are set annually for all workers, and, once established, remain fixed until the next annual negotiation and settlement.

(A2) The labor force is divided into four distinct groups on the basis of the quarter in which their annual wage negotiations and settlements take place.

(A3) The ratios of all seasonal groups in the labor force to the total labor force are constant.

(A4) The percentage change in wages for each of the four seasonal groups is a function of the same set of explanatory variables with the same parameter values for each group. Explanatory variables (X) and error term (u) are dated in the quarter in which the wage negotiation and settlement took place (j). That is,

$$\frac{w_j^h - w_{j-4}^h}{w_{j-4}^h} = aX_j + u_j$$

for $h=1, \ldots, 4$ and $j=h+4s$, where s is an integer and w_j^h is the wage rate for the hth group in the jth quarter.

(A5) The relative change in the aggregate wage rate is appropriately approximated by a moving average for the relative changes in the wage rates of the four groups. The weights of this moving average are assumed to be equal,[4] where

[4] Critical discussions of this assumption are provided by Rowley and Wilton (1972b), *Int. Econ. Rev.*, forthcoming.

w_t is the aggregate wage rate.

$$\frac{w_t - w_{t-4}}{w_{t-4}} = \sum_{h=1}^{4} 0.25 \frac{w_t^h - w_{t-4}^h}{w_{t-4}}$$

For statistical purposes, the model is completed by the specification of a list of explanatory variables and a distribution for the error term. Ignoring this latter ingredient, the final equation is the familiar one, with the overlapping annual wage change form of the dependent variable and fourth-order moving averages for the explanatory variables. However, the error term is likewise distributed as a moving average process with the same known weights.

$$\frac{w_t - w_{t-4}}{w_{t-4}} = a\left(\sum_{i=0}^{3} 0.25\, X_{t-i} \right) + \sum_{i=0}^{3} 0.25\, u_{t-i}$$

Let $\{u_t\}$ be a sequence of normally distributed errors which are mutually independent and have constant, equal variances. With this additional assumption and the knowledge of the known weights, the method of generalized least squares can be employed to calculate efficient estimates of the coefficients of the explanatory variables. As shown in our Appendix, knowledge of the weights indicates a transformation which can be used to adjust the equation so that the problem of autocorrelation is eliminated.[5] Then, the appropriate estimates of the standard errors of estimated coefficients (based on unbiased estimates of their variances) and Student's t-statistics can be calculated.

In the absence of this adjustment, the *OLS* method yields biased estimates of standard errors and invalid Student's t-statistics unless the sample fulfills the conditions given in the references cited above. Calculation of *GLS*[6] estimates permits us to observe how badly the *OLS* estimates are for the hypothetical model with a given sample. Note that even if this distribution of $\{u_t\}$ is inappropriate, as would be the case if the original errors for individual groups were autocorrelated, the results would remain of interest since they reveal the sensitivities of estimates to different specifications of the distribution for errors.[7] Finally, notice that the aggregate error is not generated by a first-order autoregressive process so the Durbin-Watson statistic is not an appropriate indicator of serial correlation and applications of the popular Hildreth-Lu procedure "to correct for serial correlation" will not do so in fact.[8]

II. Empirical Results

Results for the model of Perry are supplemented below by additional estimates for the model of Bodkin et al. and the wage equation contained in the econometric model (*RDXI*) of the Bank of Canada.[9] The role anticipated for these supplemental results stems from the danger that the inadequacies of empirical results for a single study might prove atypical. Although the use of three studies cannot eliminate this danger, we can

[5] A more detailed account is provided by Rowley and Wilton, *Metroeconomica*, forthcoming.

[6] *GLS* estimates are seldom calculated even if we have precise knowledge of the process generating the errors because of their computational burden. J. H. Wilkinson demonstrates how this burden can be substantially reduced by the efficient Cholesky technique in our particular case of a moving average.

[7] Rowley and Wilton (1972c) indicate suitable approaches in the context of alternative distributions of the errors and wage-spillover effects between different micro-groups.

[8] Empirical evidence on the loss of efficiency which stems from this procedure is presented by Rowley and Wilton (1972a).

[9] Our choice of studies was affected by our background and by the accessibility of data. In the latter regard, we wish to thank the Bank of Canada and Steve Kaliski for making data available to us. Both of the additional studies employ different wage data and time horizons from the study by Perry.

present conclusions with greater confidence than would be the case with a single study.

Each of the following three tables are divided vertically into two sections. The first section contains *OLS* estimates based on inappropriate formulae for standard error and Student's t-statistics (columns (1), (2), (3) for Table 1, columns (1), (2), (3), (4) for Table 2 and column (1) for Table 3), whereas the second section contains the appropriate *GLS* estimates on the assumption that the errors follow the distribution cited. Student's t-statistics are presented in parentheses below each estimated coefficient. To clarify exposition, the calculated numbers for the *OLS* estimates of Student's t-statistics are prefixed by the term "pseudo" since we have reason to believe that they are based

TABLE 1—PERRY WAGE EQUATIONS FOR *U.S.* MANUFACTURING INDUSTRY (1948 II–1960 III)

	Ordinary Least Squares[a]			Generalized Least Squares[b]		
	(1)	(2)	(3)	(4)	(5)	(6)
$\dot{C}_{t-1}$	0.385	0.327	0.518	0.183	0.267	0.313
	(6.946)	(5.487)	(8.717)	(1.614)	(2.553)	(2.151)
$1/U_t$	14.611	10.344	14.601	12.795	6.506	9.352
	(6.716)	(4.659)	(5.744)	(2.247)	(1.263)	(1.183)
R_{t-1}	0.434	0.524	0.223	0.544	0.521	0.283
	(6.270)	(7.058)	(3.511)	(2.954)	(3.365)	(1.437)
ΔR_t	0.832	0.714	0.527	0.856	0.559	0.664
	(4.727)	(3.529)	(3.371)	(3.065)	(1.797)	(2.395)
Const.	−4.421	−4.659	−2.171	−4.949	−3.657	−1.393
D.W.	1.188	0.811	1.358	2.19	2.32	2.06
$F(5, 44)$	71.706	62.825	59.967	8.713	11.727	4.130

Dependent variable = Annual percentage change in straight-time hourly earnings of production workers for total, durable, and nondurable manufacturing

$$[(W_t - W_{t-4})/W_{t-4}]$$

$\dot{C}_{t-1}$ = Four-quarter moving average of one-quarter percentage change in the consumer price index

$$\left(\sum_{i=0}^{3} \frac{C_{t-i} - C_{t-1-i}}{C_{t-1-i}}\right)$$

lagged one quarter.

$1/U_t$ = Reciprocal of the four-quarter moving average of the unemployment rate.

R_{t-1} = Four-quarter moving average for the annual profit rate (ratio of corporate earnings after taxes to stockholders equity), lagged one quarter, for total, durable and nondurable manufacturing.

ΔR_t = first difference of the profit rate series.

$\{(1), (4)\}$ = Total manufacturing

$\{(2), (5)\}$ = Nondurables manufacturing

$\{(3), (6)\}$ = Durables manufacturing

[a] Pseudo t-statistics are shown in parentheses.

[b] t-statistics are shown in parentheses.

TABLE 2—BODKIN ET AL. WAGE EQUATIONS FOR CANADIAN MANUFACTURING INDUSTRY (1953 I–1965 II)

	Ordinary Least Squares[a]				Generalized Least Squares[b]			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\dot{P}_t$	0.377 (4.881)	0.458 (5.354)	0.428 (5.256)	0.541 (5.995)	0.250 (1.085)	0.276 (1.186)	0.238 (1.001)	0.274 (1.117)
$1/U_t^2$	10.427 (1.553)	22.839 (3.344)	22.931 (4.070)	41.590 (10.016)	−5.514 (−0.390)	4.730 (0.381)	4.079 (0.300)	24.114 (2.354)
$(Z/Q)_{t-2}$	0.053 (2.947)	0.066 (3.292)			0.077 (1.925)	0.096 (2.499)		
$\dot{W}_{us_t}$	0.432 (3.948)		0.495 (4.261)		0.444 (1.456)		0.633 (2.131)	
$\dot{W}_{t-4}$	−0.092 (−2.279)	−0.124 (−2.740)	−0.100 (−2.306)	−0.141 (−2.851)	0.009 (0.117)	−0.021 (−0.260)	0.064 (0.821)	0.038 (0.469)
Const.	−4.122	−4.566	0.807	1.809	−6.403	−7.226	0.714	2.169
$D.W.$	1.45	1.11	1.28	0.96	2.31	2.25	2.35	2.27
F (,)	44.335	38.907	45.480	39.760	5.419	6.092	5.516	5.423

Dependent variable = Annual percentage change in average hourly earnings of production workers in Canadian manufacturing industry

$$[(W_t - W_{t-4})/W_{t-4}]$$

$\dot{P}_t$ = Four-quarter moving average of annual percentage change in consumer price index

$$\left[1/4 \sum_{i=0}^{3} \left[\frac{P_{t-i} - P_{t-4-i}}{P_{t-4-i}}\right)\right]$$

$1/U_t^2$ = Squared reciprocal of a four-quarter moving average of a two-quarter average of the Canadian unemployment rate

$$\left[1/8\, U_t + 1/4 \sum_{i=1}^{3} U_{t-i} + 1/8\, U_{t-4}\right]^{-2}$$

$(Z/Q)_{t-2}$ = Four-quarter moving average of the profit markup on output (index of corporate profits before tax divided by manufacturing production index), lagged two quarters.

$\dot{W}_{us_t}$ = Four-quarter moving average of the annual percentage change in average hourly earnings in *U.S.* manufacturing expressed in *U.S.* dollars

$$\left(1/4 \sum_{i=0}^{3} \frac{W_{us_{t-i}} - W_{us_{t-4-i}}}{W_{us_{t-4-i}}}\right)$$

$\dot{W}_{t-4}$ = Dependent variable lagged four quarters.

[a] Pseudo t-statistics all shown in parentheses.
[b] t-statistics are shown in parentheses.

on biased estimates of variances and on a misspecification of the distribution for the errors.

Two questions should be considered while the empirical results are read. First, would the set of *OLS* estimates lead the (typical) economist to draw invalid inferences with respect to the significance of variables in the models. Second, if these invalid inferences are drawn, are they sufficiently severe to suggest that the additional burden of computation for the

TABLE 3—*RDXI* WAGE EQUATION FOR CANADIAN ECONOMY (1955 I–1965 IV)

	Ordinary Least Squares[a] (1)	Generalized Least Squares[b] (2)
$\dot{P}_t$	0.940	0.516
	(3.922)	(0.815)
$1/U_t^2$	0.521	0.309
	(4.590)	(0.795)
R_t	0.867	1.104
	(2.372)	(1.072)
$\dot{W}_{t-4}$	−0.224	0.154
	(−1.624)	(1.099)
Const.	−4.712	−6.284
D.W.	1.75	2.61
$F(5, 39)$	111.805	7.343

Dependent variable = Annual percentage change in the average hourly wage in the private sector $[(W_t - W_{t-4})/W_{t-4}]$

$\dot{P}_t$ = four-quarter moving average of the annual change in the implicit price index of consumer nondurable expenditure

$$\frac{1}{4}\sum_{i=0}^{3}\frac{P_{t-i} - P_{t-4-i}}{P_{t-4-i}}$$

$1/U_t^2$ = squared reciprocal of a four-quarter moving average of the unemployment rate.

R_t = four-quarter moving average of the ratio of corporate profits before taxes (less income tax accruals) to real gross national expenditure.

$\dot{W}_{t-4}$ = dependent variable lagged four quarters.

[a] Pseudo t-statistics are shown in parentheses.

[b] t-statistics are shown in parentheses.

GLS estimates is worthwhile.

Looking only at the *OLS* estimates in these three tables, one would conclude that these Phillips curve variants perform reasonably well. The three common explanatory variables, unemployment, price increases, and profitability, are pseudo-significant in all cases save one (the unemployment variable in equation (1) of Table 2). With the possible exception of the Durbin-Watson statistic, all of the conventional statistical checks are satisfied, i.e., sign patterns, Student t-statistics greater than two, etc.

On the other hand, statistical inferences drawn from the *GLS* equations are largely inconclusive. While F-tests indicate that in all cases the joint hypothesis that all coefficients for explanatory variables are zero must be rejected (since critical .05 values for the F-statistic fall within the 2.4–3.5 range), the results for individual explanatory variables are not encouraging. Only the profit variable retains significance in even half of the *GLS* equations. Unemployment and price increases, perhaps the two most common explanatory variables in all wage research, are significant in only two of eight individual cases. One can only speculate whether the various authors would have advanced the Phillips curve model had they been faced with the *GLS* estimates rather than the *OLS* estimates.

Turning from individual explanatory variables to a comparison of the three studies, the Perry model holds up best under *GLS*, but again the results are not that encouraging. While all variables are pseudo-significant under *OLS*, there is no consistent pattern for variables under *GLS*. No variable's coefficient retains significance in each of the three sectors. The basic Phillips curve variable is significant in only one sector while price and profit variables are significant in two of three sectors. In sectoral terms, no one equation exhibits significant coefficients for all four explanatory variables.

The *GLS* results for the two Canadian studies are perhaps more conclusive. None of the usual explanatory variables would appear to be reliable from usual statistical tests. In place of fifteen pseudo-significant coefficients in the Bodkin et al. study, only three parameter estimates are

significant under *GLS*. Their basic model, as represented by equations (1) and (5) in Table 2, fails to retain one of the five pseudo-significant parameters. As shown in Table 3, all explanatory variables in the *RDXI* wage equation likewise prove to be insignificant under *GLS*, again a revealing contrast from the *OLS* results. The only consolation rests with the *F*-statistic, even though it is greatly reduced under *GLS*. As a group the explanatory variables are significant at the .05 level.

In short, the potential bias in estimated Student's *t*-statistics using conventional *OLS* techniques appears to be substantial. Employing the *GLS* technique, which produces unbiased estimates of the variances of estimated coefficients, all standard error estimates exceed those obtained using *OLS* with more than two-thirds of the individual coefficient standard errors increasing by more than 100 percent. In terms of standard significance tests, the results are even more pronounced. All Student's *t*-statistics fall under *GLS* as *OLS* pseudo-*t*-statistics are at least twice as large for over three quarters of all parameter estimates and at least three times as large for more than half of the parameter coefficients.

Two final comments can be made concerning the estimates. Since both *OLS* and *GLS* give unbiased parameter estimates, one might expect our coefficient estimates to be reasonably similar for both estimation techniques. Almost two-thirds (20 of 32) of the *OLS* parameter estimates are within one standard error of the *GLS* estimates. Finally, the appropriateness of the Durbin-Watson statistic is somewhat dubious. As pointed out above, it fails to test for a fourth-order moving-average process in the error term, as exists in the *OLS* equations. While the Durbin-Watson statistics are close to two under *GLS*, a fourth-order autoregressive pattern would appear more plausible as an alternative hypothesis in view of the annual bargaining patterns in the four distinct labor groups.

III. Conclusion

A necessary first step in empirical research is the specification of the form of the dependent variable, that is, the sequence of values which are to be explained. Quarterly analyses of wage determination have adopted overlapping annual changes as the appropriate form of their dependent variable. Explicit recognition of the statistical implications of this choice is extremely rare. We have presented a set of assumptions which are sufficient to explain this particular specification for the dependent variable. Serial correlation in the error term is a direct consequence of the aggregative procedure indicated by these assumptions. In particular, if researchers specify the same moving average for all explanatory variables, then the aggregate error must exhibit the same moving average features. Hence, all of the conventional formulae for appropriate standard errors and test statistics are potentially biased, with the extent of biases to be determined empirically.

For the three models which we have considered, these biases are shown to be substantial. A comparison of Student's *t*-statistics based on correct formulae with the pseudo-results suggest that the latter must be reduced by at least 50 percent in most cases. The inferential consequences of these reductions are severe and cannot be neglected. The roles of certain variables, which are usually considered to be of major importance in the determination of wage levels, are in doubt. For example, in the eight equations for which *GLS* estimates are calculated, unemployment is significant at the .05 level twice and not seven times as indicated by the results for the conventional *OLS* procedure. Similarly, the change in prices is significant

twice at this critical level although the *OLS* results indicate eight times.

While there are dangers in attempts to generalize these results which are based on only three studies, every test indicates the potential severity of the problem of unacknowledged serial correlation.[10] Fortunately, as we have demonstrated, there are simple ways to correct for this bias in estimation. Correctional techniques require few additional assumptions other than those which already form the bases for specifications of the dependent variable and moving averages of the explanatory variables. Only after these corrections are made will economists be able to assess the true significance of the alternative theories which have been advanced to explain the determination of aggregate wage changes.

Appendix

The collection of error terms $\{u_t\}$ from the equations for the individual labor groups is assumed to form a temporal sequence of normally distributed random variables with zero means. This sequence is assumed to be free from autocorrelation and heteroscedasticity[11] so that attention may be solely focused on the specific problem introduced by the aggregative process, which indicates a sequence of moving-average aggregative errors, $\{e_t\}$ say. Thus, from (A5), we have the following definition.

$$(1) \qquad e_t \equiv 0.25 \sum_{i=0}^{3} u_{t-i} \qquad \text{for all t}$$

Since the individual errors have zero means and the weights for aggregation are non-stochastic, the aggregative errors have zero means and their autocovariance function can be represented by mean products of the form $E(e_t e_{t-s})$ where $E(\cdot)$ represents the mean or mathematical-expectation operator. Due to our stationarity assumptions, the products will only depend on the lag s between the individual errors and can be denoted as γ_s.

$$(2) \quad \gamma_s \equiv E(e_t e_{t-s}) \qquad \text{for all } s, \text{t}$$
$$= 0.0625E\left[\left(\sum_{i=0}^{3} u_{t-i}\right)\cdot\left(\sum_{j=0}^{3} u_{t-j-s}\right)\right]$$

$$(3) \quad \gamma_s = 0.0625 \sum_{i=0}^{3}\sum_{j=0}^{3} E(u_{t-i}u_{t-j-s}) \qquad \text{for all } s, \text{t}$$

In view of the assumptions with respect to autocorrelation and homoscedasticity which were made explicit above, this expression can be simplified. Let σ^2 represent the common variance of the individual errors and we have

$$(4) \quad \gamma_s = 0.0625\,(4-s)\sigma^2 \quad \text{for } s = 0, 1, 2, 3$$

$$(5) \quad \gamma_s = 0 \quad \text{for } s > 4$$

$$(6) \quad \gamma_s = \gamma_{-s} \quad \text{for all } s$$

Thus, the autocovariances $\{\gamma_s\}$ are known apart from the scale factor σ^2 and the dispersion matrix of the aggregative errors can be constructed. This is a Laurent matrix with a narrow band of positive elements about its principal diagonal representing the Yule-Slutsky effect of the aggregative process:

$$0.0625\sigma^2 \begin{bmatrix} 4 & 3 & 2 & 1 & & & & 0 \\ 3 & 4 & 3 & 2 & 1 & & & \\ 2 & 3 & & & & & & \\ 1 & & & & & & & \\ & 1 & & & & & & 1 \\ & & & & & & & 2 \\ & & & 1 & 2 & 3 & 4 & 3 \\ 0 & & & & 1 & 2 & 3 & 4 \end{bmatrix}$$

Knowledge of this matrix (apart from the scale factor) is sufficient for the generalized least squares estimates to be calculated. See,

[10] Since this paper was originally drafted, we have reestimated other models with similar results. For example, only one of seven unemployment coefficients retains significance at the .05 level in *GLS* estimates for equations (2)–(5) in Pierson.

[11] Clearly the influences of many other factors are being ignored and these may confound the empirical results. We are, for example, ignoring the potential problems associated with errors in measurement, stochastic weights and simultaneity.

for example, the descriptions by Jack Johnston, pp. 208–11, and Arthur Goldberger, pp. 231–35. These indicate an appropriate transformation for the model which is based on a factorization of the dispersion matrix and its inverse. In practice, inversions of large matrices must be efficiently computed so that numerical errors do not distort the resultant numbers. The Cholesky technique for inversion is especially recommended in this context since it takes account of the symmetry and "bandedness" of the matrix. An account of this technique is provided by J. Wilkinson.

REFERENCES

P. S. Anderson, "Wages and the Guideposts: Comment," *Amer. Econ. Rev.*, June 1969, *59*, 351–54.

R. G. Bodkin et al., *Price Stability and High Employment: The Options for Canadian Economic Policy, An Econometric Study*, Ottawa 1967.

L. A. Dicks-Mireaux and J. C. R. Dow, "The Determinants of Wage Inflation: United Kingdom, 1946–56," *J. Royal Statist. Soc.*, 1959, A, *22*, 145–74.

G. De Menil, "Nonlinearity in a Wage Equation for United States Manufacturing," *Rev. Econ. Statist.*, May 1969, *51*, 202–06.

O. Eckstein, "Money Wage Determination Revisited," *Rev. Econ. Stud.*, Apr. 1968, *35*, 133–43.

M. L. Evans and L. R. Klein, *The Wharton Econometric Forecasting Model*, Philadelphia 1968.

A. S. Goldberger, *Econometric Theory*, New York 1963.

J. Helliwell et al., "The Structure of RDXI," *Bank of Canada Staff Research Studies*, *3*, 1969.

J. Johnston, *Econometric Methods*, 2d ed., New York 1971.

L. R. Klein and R. J. Ball, "Some Econometrics of the Determination of Absolute Prices and Wages," *Econ. J.*, Sept. 1959, *69*, 465–82.

W. Kruskal, "When are Gauss-Markov and Least Squares Estimators Identical? A Coordinate-Free Approach," *Ann. Math. Statist.*, Feb. 1968, *39*, 70–75.

E. Kuh, "A Productivity Theory of Wage Levels—An Alternative to the Phillips Curve," *Rev. Econ. Stud.*, Oct. 1967, *34*, 333–60.

M. E. Levy, "U.S. Phillips Curves and Target Unemployment Rates," *Bus. Econ.*, Fall 1967, *2*, 70–76.

R. G. Lipsey and J. M. Parkin, "Incomes Policy: A Re-Appraisal," *Economica*, May 1970, *37*, 115–38.

S. K. Mitra and C. R. Rao, "Conditions for Optimality and Validity of Simple Least Squares Theory," *Ann. Math. Statist.*, Oct. 1969, *40*, 1617–1624.

G. L. Perry, *Unemployment, Money Wage Rates, and Inflation*, Cambridge 1966.

E. S. Phelps, "Money-Wage Dynamics and Labor-Market Equilibrium," *J. Polit. Econ.*, July-Aug. 1968, *76*, 678–711.

——— **et al.**, *Microeconomic Foundations of Employment and Inflation Theory*, New York, 1970.

A. W. Phillips, "The Relation Between Unemployment and the Rate of Change of Money Wage Rates in the United Kingdom, 1861–1957," *Economica*, Nov. 1958, *25*, 283–99.

G. L. Pierson, "Union Strength and the U.S. Phillips Curve," *Amer. Econ. Rev.*, June 1968, *58*, 456–67.

C. R. Rao, "Least Squares Theory Using an Estimated Dispersion Matrix and its Applications to Measurement of Signals," in L. M. LeCam and J. Neyman, eds., *Proceedings of the Fifth Berkeley Symposium for Mathematical Statistics and Probability*, Vol. 1, 1967, 355–72.

———, "A Note on a Previous Lemma in the Theory of Least Squares and Some Further Results," *Sankhya*, Sept. 1968, A, *30*, 259–66.

G. L. Reuber, "Wage Adjustments in Canadian Industry, 1953–66," *Rev. Econ. Stud.*, Oct. 1970, *37*, 449–68.

J. C. R. Rowley and D. A. Wilton, "Wage Determination: The Use of Instrumental Assumptions," *Int. Econ. Rev.*, forthcoming.

——— **and** ———, *Implications for Estimation of Conventional Specifications in Em-*

pirical Studies of Wage-Determination, disc. pap. no. 37B, Inst. Econ. Res., Queen's Univ., Feb. 1971.

——— and ———, "Moving-Average Transformations in Classical Linear Models," *Metroeconomica,* forthcoming.

——— and ———, (1972a) *Autocorrelation in Empirical Studies of Wage Determination,* disc. pap. no. 74, Inst. Econ. Res., Queen's Univ., Mar. 1972.

——— and ———, (1972b) *The Sensitivity of Quarterly Models of Wage Determination to Aggregation Assumptions,* disc. pap. no. 79, Inst. Econ. Res., Queen's Univ., May 1972.

——— and ———, (1972c) *Temporal Spillover and Autocorrelation in Some Aggregative Models of Wage Determination,* disc. pap. no. 82, Inst. Econ. Res., Queen's Univ., May 1972.

J. D. Sargan, "Wages and Prices in the United Kingdom: A Study in Econometric Methodology," in P. E. Hart, G. Mills, and J. K. Whitaker, eds., *Econometric Analysis for National Economic Planning,* London 1964.

C. Schultze and J. Tryon, "Wages and Prices," in J. S. Duesenberry et al., eds, *Brookings Quarterly Econometric Model of the United States Economy,* Chicago 1965.

N. J. Simler, and A. Tella, "Labor Reserves and the Phillips Curve," *Rev. Econ. Statist.,* Feb. 1968, *50,* 32–49.

G. R. Sparks, and D. A. Wilton, "Determinants of Negotiated Wage Increases: An Empirical Analysis," *Econometrica,* Sept., 1971, *39,* 739–50.

B. F. Swindel, "On the Bias of Some Least-Squares Estimators of Variance in a General Linear Model," *Biometrika,* 1968, *55,* 313–16.

W. Vroman, "Manufacturing Wage Behaviour with Special Reference to the Period 1962–1966," *Rev. Econ. Statist.,* May 1970, *52,* 160–67.

G. S. Watson, "Linear Least Squares Regression," *Ann. Math. Statist.,* Dec. 1967, *38,* 1679–99.

———, "Serial Correlation in Regression Analysis I," *Biometrika,* 1955, *42,* 327–41.

J. H. Wilkinson, *The Algebraic Eigenvalue Problem,* Oxford 1965, 553–56.

G. Zyskind, "On Conditions for Equality of Best and Simple Linear Least Squares Estimators," *Ann. Math. Statist.,* 1962, *33,* 1502–03 (abstract).

———, "On Canonical Forms, Non-Negative Covariance Matrices and Best and Simple Least Squares Linear Estimators in Linear Models," *Ann. Math. Statist.,* Aug. 1967, *38,* 1092–1109.

Quit Rates Over Time: A Search and Information Approach

By Donald O. Parsons*

Recent developments suggest that important insights into the nature of the labor market can be gained by following an information and search approach.[1] A quit rate model is developed and estimated below which integrates two major informational factors. George Stigler's seminal work (1962), later systematically restated by Reuben Gronau using a more appropriate sequential decision process, assumed that workers know their alternative wage distribution but not particular wage bids. The worker's optimizing problem, then, is the search for wage bids. Armen Alchian and others have found it useful to consider the worker's process of estimating the wage distribution parameters as well, since these are not known to the worker either. Alchian hypothesizes that a worker's estimate of his alternative wage possibilities will lag behind actual conditions in a changing world. In terms of the quit rate considered here, a worker will quit less frequently than real conditions warrant during inflationary times (at least until he becomes adapted to a given rate of inflation) because he knows immediately when his own wage rises, but discovers only after some time that the wage rates of others have also risen. Optimizing models for these two distinct informational processes are derived in the next two sections and integrated into a quit probability model useful for econometric work.[2]

The parameters of the quit rate model thus derived are then estimated from monthly time-series data over the period 1959–68 for twenty-seven 3- and 4-digit manufacturing industries. The estimates appear to confirm the importance of the "real" search factors initially considered by Stigler. The vacancy rate or probability of getting a job offer has a strong impact with a very short lag time, while relative wages have the expected effect, although with longer lags and less regularity. Alchian's expectation hypothesis, however, receives little empirical support. The effect of nominal wage changes had no systematic effect on quit rates when the "real factors," vacancy rates, relative wages, etc., were controlled. In this respect, workers seem well attuned to demand conditions and there is no evidence that any wage rate illusion, whether optimally motivated or otherwise, poses a serious market problem.

I. The Optimal Job Search Strategy

In this paper, it is assumed that the worker attempts to maximize his expected income in a world where wage dispersions exist for workers of comparable skills both because information costs relative to

* Assistant professor, Ohio State University. Useful remarks and criticism by Lester Telser, H. Gregg Lewis, Belton Fleisher, and members of the labor workshops at the University of Chicago and Ohio State University are gratefully acknowledged.

[1] Beyond the seminal papers by George Stigler (1961, 1962), the works of Armen Alchian, Charles Holt, John McCall, Dale Mortensen (1970a, b), and Reuben Gronau are relevant here. The firm aspects of this process, emphasized by Holt and Edmund Phelps and beautifully integrated with the worker aspects into a complete market model by Mortensen (1970a), are largely ignored here.

[2] These two types of information gathering correspond to some degree to the extensive and intensive search dichotomy suggested by Albert Rees.

returns make perfect information uneconomical and because job transfer costs make even known higher wages sometimes uninteresting. Consider first the labor market demand conditions which will justify searching for an alternative job. Since, at each unit of search, the costs and returns to further search will be influenced by offers already received, the number of units of search will not usually be specified in advance, but will be determined sequentially. The worker must decide at every step whether incremental returns exceed incremental costs if he is to undertake what is for him the optimal amount of search activity.[3]

In determining the return to a unit of search, the worker is faced with a certain probability that the firm will not be hiring at all[4] and with a distribution of possible wage offers (which we shall assume is normal with mean $\overline{W}_A$ and variance σ_A^2) from firms which are hiring. Interpreting a "no job offer" search result as a zero wage offer, the expected return to a unit of search (R_i) will be the probability of a nonzero wage offer times the expected wage increase net of transfer costs, or

$$(1) \qquad R_i = \text{prob}\,\{\text{offer} \mid \text{search}\} \cdot E\{NWI \mid \text{offer}\}$$

The probability of a nonzero wage offer will presumably be an increasing function of job vacancies (v), say $g(v)$[5] where $\partial g(v)/\partial v > 0$. The expected net wage increase (NWI) will depend on the worker's present wage (W_0), his transfer costs (TC), and the parameters of the alternative wage distribution. In particular,

$$(2) \qquad E\{NWI \mid \text{offer}\} = \int_{W_C}^{\infty} (W_A - W_C) f(W_A)\, dw_A$$

where the critical wage $W_C \equiv W_0 + TC$, and $f(W_A)$ is the normal density function of alternative wages. Below the critical wage W_C, of course, wage offers have no economic value to the worker. When integrated, equation (2) becomes

$$(3) \qquad E\{NWI \mid \text{offer}\} = \sigma_A^2 f(W_C) - (W_C - \overline{W}_A)[1 - F(W_C)]$$

where $F(W_C)$ is the cumulative normal distribution.[6] The partial derivatives of the returns to search with respect to v, W_C, $\overline{W}_A$, and σ_A are as follows:

$$(4) \qquad \partial R_i/\partial v = \frac{\partial g(v)}{\partial v} \cdot E\{NWI \mid \text{offer}\} > 0$$

$$(5) \qquad \partial R_i/\partial W_C = \partial R_i/\partial W_0 = \partial R_i/\partial TC = -\,g(v)[1 - F(W_C)] < 0$$

$$(6) \qquad \partial R_i/\partial \overline{W}_A = g(v)[1 - F(W_C)] > 0$$

$$(7) \qquad \partial R_i/\partial \sigma_A = g(v)\sigma_A f(W_C) > 0$$

For given search cost, search by employed workers will rise when demand conditions improve and more potential employers are hiring workers.[7] Intuition suggests that this will be the most volatile component of the return structure. Search returns will also increase, as one would expect, when the worker's wage is low relative to his alternative wage. The second derivatives of equations (5) and (6) imply that the impact on search returns of a change in

[3] This rule, of course, ignores the possibility of initially decreasing costs of search.

[4] Since workers on the payroll already have processing and training investments made in them, job seekers must vie for new positions rather than compete directly with the employed. Future hiring costs will also be higher if a firm gets a reputation for handling its experienced workers poorly.

[5] Stigler (1962, p. 101) asserts the "no offer" probability will also be dependent on the ease of identifiability of the firm and other factors.

[6] An appendix containing a proof of this integration and of the partial derivatives to follow is available upon request from the author.

[7] The same rule should apply to unemployed workers as well, although total worker search efforts may increase in bad times since unemployed workers will tend to search more in general than employed workers.

wage diminishes the higher the relative wage. Equation (7) and its second derivative imply that an increase in the alternative wage dispersion will increase the returns to search,[8] with the effect strongest among workers with critical wages near the mean of the alternative wage distribution.

The total amount of search undertaken will depend, of course, on search and transfer costs as well as on the probable wage offers he will encounter. Presumably such costs will increase at an increasing rate, if only because of the geographic dispersion of employers. Below, this search cost function is approximated by a quadratic in search units. The wider the search, the less information is available on prospective employers, which again suggests rising marginal search costs. This cost function is assumed to be regenerative between periods since the worker can visit the same prospective employers repeatedly at the same cost per visit. Search costs will vary over the business cycle, since firms are likely to increase their supplies of new workers by reducing worker search and transfer costs. This can be accomplished by increasing advertising, arranging convenient interview times, paying moving expenses, etc., as well as by offering higher wages. Worker search costs per unit of search should decline when general demand conditions improve.

The worker who believes search for another job will be profitable is confronted with a choice of search techniques. He may quit as soon as he is convinced search is useful, or he may remain employed, searching after work hours or on days off, in which case he will quit only if he encounters a sufficiently favorable wage. Which option is chosen will affect the quit probability of a given worker and will also affect the lag structure of quits behind changes in market conditions. If workers quit when the decision to search is made, the lag will be in recognizing the new conditions more favorable to search. If workers remain employed, however, the quit lag will involve the search lag as well as the recognition lag.

The decision to quit to search or to remain employed is a complex decision involving the evaluation of two sequential programming problems. Both the search cost and search return functions will be altered by the decision to quit to search. The opportunity costs of time will be approximately equal since one approach sacrifices income possibilities while the other sacrifices leisure which should be approximately equal in value at the margin. The search cost function of the unemployed searcher will be lower to the extent economics of scale are possible because larger blocks of time are available. A salesman's transportation costs, to look at a comparable situation, will surely rise the more frequently he is required to return home for supplies. This consideration suggests workers will be more likely to quit to search the more geographically isolated they are from other areas of labor demand and the less flexible are their hours and work days.

Search returns will be similarly affected by the decision to quit to search. The risk factor is surely greater since the worker might well end up with a job paying less than his present wage. The stigma of being a quitter may also reduce the expected returns to a given quantity of search. In deciding which workers to hire (which implies a decision to invest in these workers), firms will attempt to increase their returns by choosing workers with a low quit propensity and may look with disfavor on unemployed quitters. One positive aspect on the returns side, which

[8] If wage dispersions have narrowed with improved communications, and transportation, as Stigler (1962, p. 98) believes, search efforts per worker should show a secular decline.

will increase the incentive to quit to search, is the likelihood that the higher wage from successful search will be captured more quickly if the worker concentrates his search by quitting.

Fortunately, scraps of empirical evidence allow us to avoid the detailed empirical analysis required to predict the proportion of workers who quit to search. Two national surveys of labor mobility seem to indicate that the great bulk of quitters search while employed and quit only if they have another job lined up. The surveys, conducted in 1955 and 1961, found that of those workers who undertook only one job change during the year and did so to "improve status," about 80 percent experienced no unemployment.[9] The quit model developed below assumes that workers quit only when a preferable job is uncovered, although the interpretation of the empirical results varies only slightly if this proposition is relaxed.

The timing of the quit decision will depend on which jobs are deemed preferable to the worker's present job. Both the nature of the firm's job offers and the specificity of movement and training in the new job option are critical. If job offers are cumulative (i.e., have no time limits), the worker will definitely quit if wages are offered above his present wage plus transfer costs, but he may not do so immediately (nor choose that particular job); instead, he may search further.

If, however, as is frequently the case, the job offer must be accepted or rejected at the time offered, the worker may reject a wage offer above his present wage plus transfer costs, if accepting the new job would seriously restrict future job search. The worker will, in fact, only accept such take-it-or-leave-it wage offers if the offered net wage plus expected future net search returns exceed the present wage plus expected future search returns at the old job, or if

$$(8) \quad W_1 + E(\text{Net Returns to Search} \mid W_1, TC_1) > W_0 + TC_0 + E(\text{Net Returns to Search} \mid W_0, TC_0)$$

The critical factor is TC_1, the new job's transfer costs, which would be a function of geographic isolation from alternative employers, among other factors. If $TC_1 = 0$, no limitations are put on future search and the worker would quit whenever $W_1 > W_0 + TC_0$. If the transfer costs are large, however, so the expected value of future job switch gains are seriously reduced, the worker may pass up a job offer he would otherwise eventually accept. Most jobs require little immediate specific investment by workers; on-the-job training accumulates over time and the worker can usually put off heavy expenditures in moving assets and family to the new location until he has searched to his satisfaction. Therefore, we assume below for simplicity that $TC_1 = 0$, and that a worker will quit immediately if he receives a wage offer greater than his present wage plus transfer costs.

The probability that a worker will quit during a given interval of time will be the probability of receiving a wage offer above the critical wage $W_C = W_0 + TC_0$. This quit probability will depend on the units of search undertaken and the probability of receiving a wage offer above W_C at each search effort. The probability, say p, of receiving a wage offer above W_C in a single drawing from the wage dispersion will be

$$(9) \qquad p = g(v)[1 - F(W_C)]$$

where $g(v)$, again, is the probability of a nonzero wage offer, a positively inclined function of labor demand v, and where $F(W_C)$ is the value of the cumulative dis-

[9] See Gertrude Bancroft and Stuart Garfinkle. In the year 1961, only 15 percent of those who experienced unemployment during the year did so because of a job quit.

tribution function at $W_C = W_0 + TC_0$. Assuming independent trials and that p is constant,[10] a worker who carries out sequential search will have the quit probability set out below,

$$\text{(10)} \quad \text{prob}\{\text{quit}\} = 1 - (1 - p)^N$$

where N is the maximum number of searches the worker will carry out if he is unsuccessful in all attempts (that is, where the unsuccessful searcher's marginal search returns (MR) equal marginal cost (MC)). This follows directly from the fact that $(1-p)^N$ is the probability of *not* quitting in N independent trials since $(1-p)$ is the probability of not quitting in any given trial. Within the period, quits will be distributed according to a geometric lag distribution.

To make the model operational, assume that the total search cost function is quadratic, so $C = bn^2$ with $\partial C/\partial n = 2bn$ where n is the number of searches. In this case, the optimal number of searches N will be,

$$\text{(11)} \quad N = \frac{g(v)}{2b} \int_{W_C}^{\infty} (W_A - W_C) f(W_A) dw_A$$

where all terms have been previously defined. The quit probability, prob $\{\text{quit}\} = 1-(1-p)^N$, behaves in a manner one might expect from the search considerations alone.[11] More job openings, a higher v, will increase the quit probabilities, as will a lower critical wage (whether due to a lower present wage, W_0, or to lower transfer costs, TC) relative to alternative wages. An increased dispersion of alternative wages will also generally increase the quit propensity, although possibly not for extremely low wage workers. The effect of job openings and relative wages will, in fact, be stronger on quit behavior than on search behavior, since these factors affect the likelihood of a quit per unit of search as well as the amount of search. A higher own wage or transfer costs will reduce the quit probability both because the probability of a quit for any amount of search is less and because the economically rational amount of search is less. When aggregated over many individuals, the quit probability equation gives the proportion of employed searchers who are likely to quit in a given cost period. This equation, therefore, is the quit rate model we seek.

Certain simplifications are required, however, before the model can be estimated conveniently. A useful approximation for the quit function can be had by expanding the quit probability function (10), using the binomial theorem,[12] to get

$$\text{(12)} \quad q = \text{prob}\{\text{quit}\} = 1 - (1 - p)^N$$
$$= Np - \frac{N(N-1)}{2} p^2$$
$$+ \text{ higher order terms in } p$$

and then approximating q by dropping the quadratic and higher order terms in p, yielding

$$\text{(13)} \quad q \cong Np$$

But $p = g(v)\phi(W_C)$

where

$$\phi(W_C) \equiv 1 - F(W_C)$$

and

$$N = \frac{g(v)}{2b} [\sigma_A^2 f(W_C) - (W_C - \overline{W}_A)\phi(W_C)]$$
$$= \frac{g(v)}{2b} \phi(W_C)\psi(W_C)$$

where

$$\psi(W_C) = \sigma_A^2 f(W_C)/\phi(W_C) - (W_C - \overline{W}_A)$$

[10] Assuming p constant assumes, in particular, that transfer costs are constant with respect to search.

[11] Formal derivations of the following assertions may be found in an appendix available from the author.

[12] This approach to approximating the theoretical model was suggested by H. Gregg Lewis.

Substituting N and p into (13) gives

$$(14)\qquad q = \frac{[g(v)]^2}{2b}\phi^2\psi = \frac{g(v)^2}{2b}\mu(W_C)$$

where $\mu=\phi^2\psi$. This suggest a partitioning of the quit factors into three parts as the *log* of (14) makes clear,

$$(15)\quad \log q = 2 \log g(v) - \log 2b + \log \mu(W_C)$$

A worker's quit propensity will be an additive function in *logs* of alternative "demand" factors, search cost factors, and wage considerations. Equation (15) will form the core of the quit rate model estimated in the empirical section, although a number of variables, particularly the vacancy rate and the parameters of the alternative wage distribution, are not known with certainty. It is to this second problem that the next section is devoted.

II. Informal Information Flows

The sequential search model developed in the previous section indicates the region of profitable job search with respect to the alternative wage distribution parameters, but the worker in an uncertain world is not given the value of these parameters. The estimation of the wage distribution presumably results from the consideration of informal information flows. The worker is continually receiving bits of information on his wage alternatives from friends and relatives, newspapers, etc., but he faces the problem of how to handle this flow. Is a higher wage paid to a friend a random sampling fluctuation or a sign that wages in other firms are generally higher than the worker's own wage? Presumably enough bits of information indicating others in the same skill class are receiving higher wages (or more precisely, wages that will make search profitable) will convince the worker that his wage is out of line.

An optimal technique which accords well with intuitive notions of how workers handle such information is Abraham Wald's sequential probability ratio test. In the case of a simple hypothesis, the ratio test is a likelihood test in which observations are taken until the ratio of the probability of the sample under the alternative hypothesis to that under the null hypothesis exceeds some constant A (at which point the null hypothesis is rejected) or falls below some constant B (at which point the null hypothesis is accepted). Obviously in this case, a one tail-test is appropriate since the worker is only interested in knowing whether the wage population mean exceeds some value. The constant A and B are functions of the desired significance and power of the test. In setting the power and significance, the worker must consider the tradeoff between the cost of assuming the alternative wages are high enough to justify search when in fact they are not (which will involve uneconomical search) and the cost of assuming the wage is not high enough to justify search when indeed it is (thus foregoing a possible wage improvement, for a time).

As Wald and Wolfowitz have demonstrated,[13] the sequential probability ratio test (which can be simply extended into a generalized Bayesian decision rule)[14] has strong optimal properties. In the case of a simple hypothesis against a simple alternative, this test will minimize expected sample size for given errors whichever hypothesis actually holds. With a normally distributed variate, the average recognition lag of an interesting parameter change is apparently about one-half that of the fixed sample approach. The more the parameter is out of line, of course, the more quickly the change will be discovered (for example, the fewer the observations that will be required to reject the null hypothesis).

[13] See G. Barrie Wetherill, p. 22.
[14] See Thomas Ferguson.

Of particular interest here is the nature of the detection lag. Although Wald demonstrated that under very restrictive assumptions, the distribution of lag time until discovery will be asymptotically normal, generally the distribution will be quite complex and skewed. Two Monte Carlo studies, using normally distributed random variables, found the distribution of discovery times to be approximately lognormal.[15] Since the normal distribution can be approximated discretely by a binomial with equal probability of success and failure, the worker's prediction of his average alternative wage (W_A^P) can be estimated by

$$(16)\quad W_{At}^P = (\overline{W}_{At})^{B_0}(\overline{W}_{At-1})^{B_1}\ldots\cdot(\overline{W}_{At-m})^{Bm}$$

where the B_i are binomial coefficients. Similar lag structures presumably exist for the other unknown variables.

The estimation of the lag structure of (16) is made difficult since both the worker's own wage and his prediction of alternative wages enter the quit function and the two are generally extremely highly correlated. The quit rate function, however, suggests that the difference of the two wage rates is the crucial variable, which means, to a logarithmic approximation, that $ln\ W_0 - ln\ W_A^P = ln\ W_0/W_A^P$ is the appropriate variable, so the two wage rates need not be entered separately.[16] Since the regression requires observable variables, (16) may be introduced to yield

$$(17)\quad \frac{W_{0t}}{W_{At}^P} = \left(\frac{W_{0t}}{\overline{W}_{At}}\right)^{B_0}\left(\frac{W_{0t-1}}{\overline{W}_{At-1}}\right)^{B_1}\ldots\cdot\left(\frac{W_{0t-m}}{\overline{W}_{At-m}}\right)^{B_m}\cdot W_{0t}^{1-B_0}\cdot W_{0t-1}^{-B_1}\ldots W_{0t-m}^{-B_m}$$

The second half of (17), a weighted average of past own wages, is more conveniently estimated in logarithmic first differences, so

$$\begin{aligned}(18)\quad & ln\,(W_{0t}^{1-B_0}\cdot W_{0t-1}^{-B_1}\ldots W_{0t-m}^{-B_m})\\ &= (1-B_0)(ln\ W_{0t} - ln\ W_{0t-1})\\ &+(1-B_0-B_1)(ln\ W_{0t-1}-ln\ W_{0t-2})\\ &+(1-B_0-B_1-B_2)\\ &\cdot(ln\ W_{0t-2}-ln\ W_{0t-3})\\ &+\ldots\end{aligned}$$

Empirically this expectational hypothesis implies that, in the quit rate regressions, the sign of the own wage change coefficients should be negative (since W_0/W_A negatively affects quits) and diminishing in absolute value with time. Intuitively, holding real wage differentials and other demand factors constant, the worker will quit less frequently when wages in general are rising because the worker will see his own wage rising before he discovers other wages have risen as well. This is essentially the Alchian hypothesis.

III. The Quit Rate Regression Model and Results

The basic quit rate model to be estimated here is equation (16),

$$log\ q = 2\ log\ g(v) - log\ 2b + log\ \mu(W_C, \overline{W}_A, \sigma)$$

with the expectational lag structure for alternative wages and other unknown variables specified as in the previous section. This complete quit rate model is estimated (with certain modifications discussed below) using monthly time-series data for twenty-seven disaggregated manufacturing industries over the period January 1959, to December 1968. The twenty-seven industries, while by no means com-

[15] A summary of these is reported in Wetherill, pp. 21–22.

[16] The wage prediction problem is compounded by the fact that, in the real world, future wages are not stationary so the worker must attempt to forecast future wages paid by his own and other firms as well as predict prevailing wages.

prehensive, nevertheless cover a broad range of industry types from heavily concentrated and unionized durable goods industries to highly decentralized consumer goods industries.[17]

The quit rate data and wage rate measures are derived from Bureau of Labor Statistics publications. The all-manufacturing average wage rate was used as a proxy for the mean of the alternative wage distribution ($\overline{W}_A$). For the change in nominal own wage, important in the expectational hypothesis, industry average hourly wages were corrected for changes in overtime pay where possible to give a measure of straight time wage changes.[18]

The vacancy rate data used here are derived from job offers pending with the Bureau of Employment Security, end of month.[19] This measure is economy wide rather than an industry specific job vacancy rate. Although both types of measures would be desirable, only the former is available. This measure should be the more useful of the two in any case since individual surveys indicate most workers do shift industries when they quit. Changes in industry output was included to capture the pure industry labor demand. The interpretation of the output change coefficients, however, is made difficult by auxiliary factors discussed below.

Several other factors suggested by the quit equation as appropriate in the regression analysis lack direct measures and are either ignored or handled in a rather *ad hoc* fashion.[20] In the latter category are job transfer costs, an important element in W_C, the critical wage. Transfer costs will vary both cyclically (when firms become more willing to absorb such costs) and seasonally. The sudden jump in quit rates in late summer, for example, is the result of family heads moving before the school year begins as well as school children quitting work to return to school. An August-September dummy was introduced to capture this effect.

Another major determinant of average worker job transfer costs will be the relative inflow of new workers into the industry. The initial investment a new worker has to make in his job need not be large. Not only can transfer costs be avoided until the worker is sure he wants this particular job, but also the specific human capital arising from "on-the-job training" requires time to accumulate. Time-series seniority data by industry are not available, but relative change in output should be positively correlated with the proportion of workers with little seniority.

The lag structures of relative wages and change in own wage, as suggested by (17) and (18) were estimated in the following ways. The change in own wage rate was entered with a number of lagged values, since the change variables were not highly correlated, and the appropriate number of lagged variables was determined empirically by consideration of experimental regression on three randomly selected industries (336, 365, 324). Similar methods were used with the change in *log* output variables. In each, two lagged monthly values were included. For the relative wage variable, a six period geometric average seemed to fit best in the experimental regressions,

[17] This subset of manufacturing industries was selected because of the availability of monthly output measures independent of employment figures, as reported by Hamermesh.

[18] This correction was possible for all 3-digit industries in the sample but not 4-digit industries since overtime hours are reported only for the more aggregated classifications.

[19] The Bureau of Employment Security figures were kindly supplied to me by J. Peter Mattila. They were compiled by the National Bureau of Economic Research.

[20] In the former category are the alternative wage dispersion and direct search costs. Crude evidence suggests that wage dispersions narrow during prosperity while search costs are likely to fall since firms are more willing to advertise, etc., when the labor market is tight

although the binomially weighted average did no better than the uniformly weighted average so the latter was included in the final regressions. Finally, the vacancy rate lag structure also seemed to fit best with three months of vacancy measures. These were introduced individually and unconstrained although multicollinearity problems cannot be ignored here.

The ultimate functional form for the regression model used to estimate the quit rate function for the twenty-seven industries was the *log* of q in equation (19).

$$q = e^{\alpha_0+\alpha_1 AS} \cdot v^{\alpha_2} \cdot v_{-1}^{\alpha_3} \cdot v_{-2}^{\alpha_4} \tag{19}$$

$$\cdot \left(\frac{W_0}{W_A}\right)_6^{\alpha_5} \cdot \left(\frac{W_0}{W_{0_{-1}}}\right)^{\alpha_6} \cdot \left(\frac{W_{0_{-1}}}{W_{0_{-2}}}\right)^{\alpha_7} \cdot \left(\frac{W_{0_{-2}}}{W_{0_{-3}}}\right)^{\alpha_8}$$

$$\cdot \left(\frac{O}{O_{-1}}\right)^{\alpha_9} \cdot \left(\frac{O_{-1}}{O_{-2}}\right)^{\alpha_{10}} \cdot \left(\frac{O_{-2}}{O_{-3}}\right)^{\alpha_{11}}$$

The results for the individual industries appear in Table 1 for all the variable coefficients except the output variables which are reported in an appendix available from the author upon request. The regressions were corrected for first-order autocorrelation as noted in Table 1 when the null hypothesis of zero autocorrelation was not accepted at the 1 percent level, using the Durbin-Watson statistic.

The vacancy rate coefficients suggest that job openings are a major determinant of month-to-month variation in industrial quit rate performance. At least one of these coefficients is significantly positive at the 0.95 confidence level in twenty-four of twenty-seven industries and normally strongly so, with modal quit rate-vacancy rate elasticities usually in the range 1.0–2.0. The impact of job openings is substantial in the first period in twenty-two of twenty-seven industries with the major vacancy rate impact in the remaining industries occurring in the second month.[21] By the third month, the influence is much weakened and more likely to be perverse in sign. There is a mild, but insignificant, tendency for the modal quit elasticity to be somewhat higher in industries with high average production worker wages.

The six month geometric average of relative wages has a less uniformly strong effect on industry quit rates. The relative wage coefficient is negative in seventeen of twenty-seven industries and significantly negative in nine of twenty-seven. The estimated quit rate elasticities, however, tend to be larger in absolute terms than one when negative, suggesting that wage factors can play a substantial role in some industries. The reasons for the more erratic relative wage results may well be in the limitations of the data. Not only are average production worker wage rates a measure of skill mix as well as wage scales, but the all-manufacturing average production worker wage is perhaps a poor proxy for the mean of alternative wages facing workers in any particular industry.

A result of particular interest is the fact that changes in a worker's own wage, with vacancies and relative wages fixed, has no systematic effect on quit behavior. The coefficients are more likely to be positive than negative and are rarely significant. Apparently workers learn of changes in alternative wage possibilities quickly enough that they are not fooled into believing that their own wages have risen relative to others when wages are rising in general.[22] This evidence casts some doubt on mechanisms proposed for reconciling Phillips curve phenomena in a price theory frame-

[21] Substituting unemployment rates for vacancy rates in the regressions has only minor influence on the results. The unemployment rate coefficients are generally less significant and are more likely to peak in the second or third period.

[22] Replacing changes in own wage with changes in the general price level yields comparably uninteresting results.

TABLE 1—*LOG* QUIT RATE REGRESSION COEFFICIENTS FOR SELECTED MANUFACTURING INDUSTRIES, 1959–1968, MONTHLY[a]

(*t*-values in italics)

Industry	Constant	AS[b]	LV[c]	LV_{-1}	LV_{-2}	LW_6[d]	CLW[e]	CLW_{-1}	CLW_{-2}	$\bar{R}^2$
Sawmills & Planing Mills[f]	−4.97	0.46	0.83	0.84	−0.59	0.31	1.83	2.55	0.79	0.89
242	*−13.85*	*11.54*	*3.10*	*2.13*	*−2.66*	*0.53*	*2.65*	*3.11*	*1.16*	
Millwork[f]	−4.68	0.59	1.19		−0.27	−2.33	3.45	8.15	7.90	0.86
2431	*−11.07*	*13.57*	*6.38*		*−1.62*	*−1.36*	*1.36*	*2.89*	*3.27*	
Veneer & Plywood[f]	−4.27	0.44	1.14	0.96	−1.16	0.92	0.33	0.48	−0.10	0.87
2432	*−11.04*	*11.46*	*4.90*	*2.59*	*−5.83*	*0.81*	*0.61*	*0.68*	*−0.19*	
Glass Containers[f]	−6.28	0.60	1.46		−0.23	−6.67	−1.49	−0.60	−0.29	0.79
3221	*−13.25*	*11.06*	*8.56*		*−1.38*	*−2.53*	*−1.29*	*−0.45*	*−0.28*	
Brick & Structural Clay Tile[f] 3251	−8.21	0.48	0.71	0.72		−4.48	2.63	5.37	3.17	0.87
	−19.37	*8.53*	*1.81*	*2.01*		*−2.74*	*1.02*	*2.15*	*1.42*	
Iron & Steel Foundries	−8.86	0.29	2.07	−1.15	0.54	4.11	−0.88	1.21	2.60	0.93
332[f,g]	*−28.20*	*6.08*	*8.13*	*−2.92*	*2.46*	*4.84*	*−0.52*	*0.61*	*0.14*	
Aluminum Extruding[f]	−5.33	0.61	0.99	0.54	−0.51	−4.60	−2.70	−0.44		0.74
3352	*−4.74*	*8.05*	*2.14*	*0.76*	*−1.24*	*−3.16*	*−1.79*	*−0.29*		
Nonferrous Foundries[f]	−7.21	0.25	2.48	−1.96	0.91	−1.22	−0.56	1.81	2.56	0.91
336	*−12.01*	*5.57*	*9.72*	*−4.73*	*3.79*	*−1.32*	*−0.21*	*0.61*	*0.94*	
Metal Cans[f]	−4.41	1.00	−0.70	2.44	−0.84	−1.27	−1.50	−1.69	−0.66	0.76
341	*−5.90*	*11.12*	*−1.33*	*2.96*	*−1.71*	*−1.50*	*−0.55*	*−0.50*	*−0.24*	
Heating Equipment	−5.98	0.41	1.41	−0.68	0.41	−4.07	1.70	3.08	−1.92	0.92
3433	*−12.86*	*8.54*	*5.42*	*−1.53*	*1.50*	*−2.90*	*0.78*	*1.37*	*−0.89*	
Household Refrigerators[f]	−10.74	0.33	2.19	−1.38	1.02	3.06	0.57	−5.09	−10.98	0.69
3632	*−10.04*	*2.97*	*3.33*	*−1.28*	*1.73*	*0.78*	*0.17*	*−1.41*	*−3.28*	
Household Laundry[f]	−8.56	0.68	1.50	−1.00	0.99	0.31		0.30	−0.14	0.71
3633	*−11.00*	*7.34*	*3.03*	*−1.23*	*2.03*	*0.16*		*0.69*	*−0.33*	
Radio & T.V. Receiving Sets[f] 365	−4.98	0.38	0.70	0.25	0.06	0.36	3.45	3.83	1.96	0.70
	−6.36	*6.67*	*2.14*	*0.50*	*0.20*	*0.63*	*1.91*	*1.74*	*1.08*	
Confectionary[f]	−3.61	0.44	−0.18	0.42	0.32	−3.06	0.66	−1.38	−3.72	0.81
207	*−10.21*	*5.09*	*−0.58*	*0.84*	*1.16*	*−3.91*	*0.39*	*−0.75*	*−3.02*	
Cigarettes[f]	−5.68	1.22	−0.27	1.86	−0.75	−0.80	2.83	−2.62	−2.95	0.68
211	*−3.74*	*11.17*	*−0.39*	*1.75*	*−1.26*	*−0.67*	*0.89*	*−0.76*	*−0.90*	
Cigars[f]	−4.48	0.21	0.90	−0.46	0.34	−1.11	−0.50	−2.04	−1.79	0.66
212	*−9.22*	*3.12*	*2.20*	*−0.70*	*0.84*	*−1.49*	*−0.25*	*−0.94*	*−0.89*	
Wool Weaving & Finishing 223	−4.78	0.41	1.78	−0.63	−0.25	−0.61	−1.19	−0.84	1.34	0.88
	−15.77	*7.27*	*5.92*	*−1.28*	*−0.96*	*−1.51*	*−0.48*	*−0.35*	*0.59*	
Men & Boys Suits	−1.72	0.23	1.73	−1.60	0.33	0.47	−1.94	2.28	3.73	0.76
321	*−6.08*	*5.46*	*6.76*	*−3.60*	*1.30*	*8.29*	*−1.55*	*1.72*	*3.22*	
Men & Boys Furnishings	−2.54	0.26	1.56	−0.83	0.17	1.42	0.45	1.87	−2.82	0.80
232	*−10.15*	*5.74*	*5.32*	*−1.76*	*0.59*	*3.68*	*0.35*	*1.46*	*−2.26*	
Paperboard[f]	−4.63	0.90	0.26	1.49	−1.03	1.31	3.34	2.55	2.38	0.82
263	*−5.78*	*13.19*	*0.70*	*2.46*	*−2.68*	*1.77*	*1.21*	*0.87*	*0.66*	
Plastic Materials & Resins 2821	−4.06	0.84	0.58	0.47	−0.20	−7.22	−7.14	−4.74	1.15	0.86
	−3.75	*13.27*	*1.48*	*0.74*	*−0.55*	*−2.41*	*−2.09*	*−1.31*	*0.35*	

[a] Dependent variable is natural *log* of quit rate. Quit rate = monthly quits per 100 workers, Bureau of Labor Statistics.
[b] August-September dummy.
[c] *Log* of average of nonagricultural jobs pending, end of month, for present and preceding month, Bureau of Employment Security.
[d] *Log* of six month geometric average of industry hourly wage to all manufacturing hourly wage, Bureau of Labor Statistics.
[e] Monthly change in *log* of industry wage.
[f] Corrected for first degree autocorrelation.
[g] Regression for period 1960–68.

TABLE 1—(*Continued*)

Industry	Constant	AS^b	LV^c	LV_{-1}	LW_{-2}	LW_6^d	CLW^e	CLW_{-1}	CLW_{-2}	$\bar{R}^2$
Synthetic Fibers[f]	−5.85	0.63	0.26	1.31	−0.60	−6.48	1.86	−4.35	−3.86	0.80
2823, 24	*−12.53*	*11.18*	*0.97*	*3.17*	*−2.42*	*−1.97*	*0.83*	*−1.77*	*−1.72*	
Tires & Inner Tubes[f]	−9.56	0.64	2.36	−1.65	0.94	−0.70	5.32	−0.55	−1.80	0.83
301	*−14.17*	*7.80*	*5.16*	*−2.09*	*1.96*	*−1.35*	*2.04*	*−0.22*	*−0.69*	
Rubber Products[f]	−4.95	0.35	1.21	−0.47	0.18	−2.46	−1.54	5.55	−1.39	0.88
302, 3, 6	*−10.80*	*8.05*	*4.88*	*−1.16*	*0.74*	*−2.97*	*−0.58*	*2.05*	*−0.54*	
Leather, Tanning &	−8.93	0.38	1.71	−0.62	0.54	−0.85	2.05		1.12	0.86
Finishing[f] 311	*−21.97*	*6.36*	*4.90*	*−1.12*	*1.77*	*−0.48*	*0.59*		*0.32*	
Footwear except Rubber	−3.94	0.27	1.04	−0.03	−0.02	0.92	0.90	−0.28	0.06	0.91
314	*−18.10*	*8.54*	*4.99*	*−0.08*	*−0.08*	*2.27*	*1.33*	*−0.37*	*0.09*	
Cement, Hydraulic	−5.67	1.26	−0.56	3.15	−1.61	−1.81	0.57	−0.40	−0.27	0.84
324	*−11.78*	*17.10*	*−1.03*	*3.45*	*−2.81*	*−4.46*	*0.96*	*−0.60*	*−0.47*	

work.[23] This bit of evidence, of course, is hardly conclusive. Variable definition and econometric problems make these assertions only tentative. Simultaneity problems, for example, exist here since high quit rates may well lead to wage increases as firms try to hold their workers. Nonetheless, the evidence does suggest that wage "illusion" is not an extremely strong effect.

Turning to the noninformational factors, the August-September dummy has coefficients significantly positive in all industries, documenting the importance of the vacation-school year effect on the workings of the *U.S.* economy. The output change variables (not reported here) have a strong and consistently positive effect on quit rates, presumably because of the direct industry-specific labor demand effect and because of the indirect effect from increasing the proportion of young, quit-prone workers as employment expands. The lag structure of output change is reasonably regular with peak effects in the first or, more likely, second period with quit rate effects in the third period still consequential.

In sum, over the ten-year period 1959–68, the variation in quit rate experiences for a large set of industries can be rather well explained by the variation of a small set of largely search-informational variables. The fluctuation in job openings, industry demand levels, and season of year are particularly strong. Relative wages are somewhat less so, while fluctuations in nominal wage levels generally have no particular relationship to quit rate fluctuations. Perhaps not surprisingly, "real" factors seem to dominate the movement of quit rates.

[23] An anonymous referee points out that the bulk of unemployed are new entrants and laid off workers, not quitters, so the Phillips relationship is not being directly examined. To the extent new entrants and laid off workers receive job condition information at greater cost than do employed workers, this may be an important qualification. Firms themselves may be subject to expectation lags so that vacancy rates will be unrealistically high during periods of unexpected inflation, thereby inducing quits and presumably pushing up wages.

REFERENCES

A. A. Alchian, "Information Costs, Pricing, and Resource Unemployment," in E. S. Phelps et al, eds., *Microeconomic Foundations of Employment and Inflation Theory*, New York 1970, 27–52.

G. Bancroft and S. Garfinkle, "Job Mobility in 1961," *Mon. Lab. Rev.*, Aug. 1963, *85*, 897–906.

J. F. Burton and J. E. Parker, "Voluntary Labor Mobility in the U.S. Manufacturing Sector," *Ind. Relat. Res. Assn. Proc.*, Dec. 1967, 61–70.

R. V. Eagly, "Market Power and an Interven-

ing Mechanism in Phillips Curve Analysis," *Economica*, Feb. 1965, *32*, 48–64.

T. S. Ferguson, *Mathematical Statistics: A Decision Theoretic Approach*, New York 1967.

R. Gronau, "Information and Frictional Unemployment," *Amer. Econ. Rev.*, June 1971, *61*, 290–301.

D. S. Hamermesh, "A Disaggregative Econometric Model of Gross Changes in Employment," *Yale Econ. Essays*, fall 1969, *9*, 107–46.

C. C. Holt, "Job Search, Phillips' Wage Relation, and Union Influence: Theory and Evidence," in E. S. Phelps et al., eds., *Microeconomic Foundations*, New York 1970, 53–123.

N. L. Johnson, "Sequential Analysis: A Survey," *J. Royal Stat. Soc.*, 1961, *124*, 372–411.

J. J. McCall, "Economics of Information and Job Search," *Quart. J. Econ.*, Feb. 1970, *84*, 113–26.

J. MacQueen and R. G. Miller, Jr., "Optimal Persistence Policies," *Oper. Res.*, May 1960, *8*, 362–80.

J. P. Mattila, "Quit Behavior in the Labor Market," unpublished 1970.

D. T. Mortensen, (1970a) "A Theory of Wage and Employment Dynamics," in E. S. Phelps et al., eds., *Microeconomic Foundations*, New York 1970, 167–211.

———, (1970b) "Job Search, the Duration of Unemployment, and the Phillips Curve," *Amer. Econ. Rev.*, Dec. 1970, *60*, 847–62.

E. S. Phelps, "Money Wage Dynamics and Labor Market Equilibrium," in his *Microeconomic Foundations*, New York 1970, 124–166.

M. W. Reder, "The Theory of Frictional Unemployment," *Economica*, Feb. 1969, *36*, 1–28.

A. Rees, "Information Networks in Labor Markets, *Amer. Econ. Rev. Proc.*, May 1966, *56*, 559–66.

A. M. Ross, "A New Industrial Feudalism," *Amer. Econ. Rev.*, Dec. 1958, *48*, 903–20.

G. J. Stigler, "The Economics of Information," *J. Polit. Econ.*, June 1961, *69*, 312–25.

———, "Information in the Labor Market," *J. Polit. Econ.*, Oct. 1962, supp., *70*, 94–105.

A. Wald, *Sequential Analysis*, New York 1947.

G. B. Wetherill, *Sequential Methods in Statistics*, London 1966.

Board of Governors of the Federal Reserve System, *Industrial Production Indexes*, 1959–68.

Bureau of Labor Statistics, *Employment and Earnings Statistics for the United States*, various years.

Bureau of Employment Security, unpublished data, supplied by the Nat. Bur. Econ. Res.

Employment Impacts of Textile Imports and Investment: A Vintage-Capital Model

By PETER ISARD*

"Development can be promoted by aid, but aid cannot and should not be relied on to do the whole job. The low-income countries need . . . improved access for their products to the massive markets of the industrialized nations. Such export increases must come largely in manufactured goods."

President Richard M. Nixon

Although valuable policy guidance continues to emerge from the efforts of development planners to tackle the internal problems of allocating resources within the less developed countries (*LDC*), a strong consensus of opinion maintains that external barriers are among the major obstacles to economic growth. Consequently, there is a need for complementary planning efforts which deal with the problem of shifting production in high-income countries away from the manufactures that are relatively advantageous for *LDC*s to produce for export. This latter problem is loaded with political difficulties, insofar as the large potential markets for *LDC* manufacturing exports also represent the vested interests of large and politically powerful groups within the industrialized nations. Yet there is good reason to conjecture that trade restrictions against *LDC* products would be less severe if politicians were presented with sophisticated economic estimates of the quantities of *LDC* manufactures which industrial countries can import at given short-run costs to their vested interest groups.[1] And as a challenge to the philosophical position of most economists, the time has come to outline quantitative industrial restructuring plans which support the widely preached proposition that trade liberalization can make everybody better off in the long run.

The purpose of this paper is to move toward a plan for limiting or gradually contracting the size of the *U.S.* textile sector, as measured by employment. By focusing on the economic constraints which relate textile employment, imports and investment, it is possible to compute reasonable estimates of the maximum growth of imports that is consistent with any given rate of employment change. The design of efficient policies for achieving desired employment levels, or other domestic textile objectives, can increase the development assistance provided to *LDC*s through *U.S.* textile imports.

It should be noted that there is a considerable difference of opinion on the wisdom of choosing to promote *LDC* exports of textile manufactures. Existing theories and statistical indicators of comparative advantage, both neofactor-proportions and neotechnological accounts, characterize textiles as a manufacture which is strongly favorable to *LDC* production and export.[2] But in a dynamic context, pursuit of high per capita income is an effort to shift

* International Finance Division, Federal Reserve Board. This paper is based upon my unpublished doctoral dissertation and was not influenced by anyone on the Federal Reserve staff. I am particularly grateful for guidance received from Alan S. Manne and Donald B. Keesing.

[1] For example, between 1961 and 1968, *U.S.* restrictions against textile imports not only protected vested interests, but also encouraged a 10.2 percent increase in the number of domestic textile workers. Economic foresight might have promoted a more liberal import policy in order to have avoided this "undesirable" expansion of vested interests.

[2] See Gary Hufbauer.

comparative advantage into activities in which labor productivity is high. Proponents of an outward-looking development strategy stress the importance of attempts to improve the productive and management skills of *LDC* human resources through contact and close competition with the advanced countries.[3] Thus, rather than promoting those exports which are comparatively cheap to produce under initial factor endowments and technological considerations, it is desirable for *LDC*s to engage in those export activities which, through associated learning experiences, are likely to increase most rapidly the productivity of their human resources. From this long-run viewpoint, it is questionable whether strong promotion of *LDC* textile exports is desirable.

From the viewpoint of potential short-run gains, however, the case is more conclusive. Textile and apparel products represent almost 50 percent of the manufactures which *LDC*s have managed to sell to the advanced countries (1965 data),[4] while imports from all sources in 1970 supplied less than 5 percent of the large *U.S.* market for textiles. Consequently, a long range *U.S.* plan to allow the maximum import growth consistent with a gradual contraction of the *U.S.* textile sector could provide sizeable development assistance benefits to the *LDC*s. And under complementary measures designed to neutralize the employment impacts and other potential costs to *U.S.* vested interests, the theory of international trade suggests that the United States as well as the *LDC*s could benefit from such contraction.[5]

[3] See Donald Keesing.

[4] See Hal Lary, Table 16, p. 110.

[5] The international tensions which *U.S.* textile policy has created, both in coercing limitations on exports of noncotton textiles from Japan and other East Asian countries to the United States, and during meetings of the *GATT* Cotton Textiles Committee, provide all the more reason for reconciling in a Pareto-efficient manner the foreign and domestic interests concerned.

This paper takes the following approach to characterizing an efficient *U.S.* textile policy. Section I develops a vintage-capital production model of the *U.S.* textile sector and compares several econometric estimates of the rates of embodied and disembodied technical change. These estimates suggest that the employment impact of a given volume of imports is quite sensitive to the extent of new investment taking place. On the assumption that technical change in labor-output ratios is at least two-thirds embodied, the estimates suggest that employment per unit output with new plant and equipment is between 35 and 49 percent of the unit labor requirement with marginally profitable plant and equipment. It follows that between 51 and 65 percent of the employment impact of imports can be neutralized if a cutback in new plant and equipment investment allows imports to substitute for the additional output of modern mills, rather than the output of marginally profitable mills. Obversely, a cutback in modernizing investment can allow a target employment level to be achieved with more than twice the volume of imports that would be acceptable without the investment cutback.

Section II projects several future time-paths of employment and imports that are jointly feasible under an investment-retirement policy designed to hold approximately constant the estimated age of the oldest operating plant and equipment. This describes several of the "wiser" choices that are available to *U.S.* policy makers. The target retirement age reflects a conjecture on the "maximum practical lifespan" of textile plant and equipment. No doubt, there is a moderate range for error in this conjecture, as well as in the econometric estimates; and in conclusion, Section III discusses the limitations of the analysis. Fortunately, policy can be continually adjusted from year to year in ac-

cordance with the feedback of empirical information; so initial policy formation does not require an extremely high degree of confidence in the estimated implications of each alternative.

I. A Vintage Model of the *U.S.* Textile Sector

The theoretical implications of vintage-capital production models have been developed by Leif Johansen, Robert Solow (1962), and Solow, James Tobin, Christian von Weizsacker, and Menahem Yaari. Subsequently, vintage models have been applied empirically by Richard Attiyeh and James Barr. These models are able: (i) to reconcile fixed capital-labor proportions in existing microeconomic production units with observed capital-labor substitution within industries over time; (ii) to explain the phenomenon of new investment in an industry with idle capacity; and (iii) to distinguish the industry average labor-output ratio from the unit labor coefficient on marginally profitable capital, thereby allowing more accurate estimation of the employment loss which results when imports displace the output of marginally profitable firms.

The following notation is used below.

- t = number of years after 1900
- v = the index of capital vintage of year $1900+v$
- Q_t = aggregate textile production during year t
- L_t = total textile production-worker man-hours employed during year t
- I_v = investment in vintage v plant and equipment, or gross investment during year v
- $I_{v,t}$ = vintage v capital utilized during year t
- $a_{v,t}$ = the labor-output ratio on vintage v capital during year t
- $b_{v,t}$ = the capital-output ratio on vintage v capital during year t
- $m(t)$ = the oldest (marginally profitable) vintage of capital utilized during year t
- d = the physical depreciation rate on capital, per year

It is assumed that:

(i) All investment expenditures during year v are used to purchase plant and equipment of vintage v.

(ii) There is a gestation lag of one year, such that all capital purchased during year v is operable at full capacity during year $v+1$, but cannot operate before year $v+1$.[6]

(iii) Once capital has gestated, it begins to depreciate at a constant exponential rate of d per year.

(iv) During any year in which the level of production falls short of maximum potential output, it is the older capital which stands idle while the more modern plant and equipment produces. More precisely, if any available vintage v capital is left idle during year t, all available vintage $v-1$ capital is also left idle during year t.[7]

Together, assumptions (i)–(iv) define the equations from which the vintage technical parameters can be estimated:

$$(1) \quad I_{v,t} = \begin{cases} 0 & \text{for } t-1<v \\ e^{-d(t-1-v)} I_v & \text{for } m(t)<v\leq t-1 \\ 0 & \text{for } v<m(t) \end{cases}$$

[6] The data restrict attention to annual time periods, with investment data representing gross expenditures on new plant and equipment combined. Hence there are two reasons why a gestation lag of one year seems appropriate. First, investment expenditures, and outlays for new plant in particular, may partly occur before operation can begin. And second, not all plant and equipment is installed on the first day of a given year: most new capacity is only available during a fraction of the year.

[7] This assumption is considerably stronger than it may sound insofar as it requires that whenever any available vintage v equipment is left idle during year t, all available vintage v plant is also left idle. Disaggregated data on plant and equipment investments would make possible a more appealing formulation.

with $0 < I_{m(t),t} \leq e^{-d(t-1-m(t))} I_{m(t)}$

$$(2) \qquad Q_t = \sum_{v=m(t)}^{t-1} \frac{I_{v,t}}{b_{v,t}}$$

$$(3) \qquad L_t = \sum_{v=m(t)}^{t-1} \frac{a_{v,t}}{b_{v,t}} I_{v,t}$$

Technical change is assumed to occur in the form:

$$(4) \qquad a_{v,t} = ae^{-hv}e^{-ct}$$

$$(5) \qquad b_{v,t} = be^{-kv}$$

The unknown parameters h and k represent the rates at which technical change is embodied in successive vintages of capital, while c is the rate of disembodied decline in labor-output ratios over time. There is assumed to be no disembodied change in capital-output ratios.

Equations (1)–(5) reduce to the model:

$$(6) \quad Q_t = \sum_{v=m(t)+1}^{t-1} \frac{e^{kv}}{b} e^{-d(t-1-v)} I_v + \frac{e^{km(t)}}{b} I_{m(t),t}$$

$$(7) \quad L_t = \frac{ae^{-ct}}{b} \left\{ \sum_{v=m(t)+1}^{t-1} e^{(k-h)v} e^{-d(t-1-v)} I_v + e^{(k-h)m(t)} I_{m(t),t} \right\}$$

for some m(t) and $I_{m(t),t}$ such that

$$0 < I_{m(t),t} \leq e^{-d(t-1-v)} I_{m(t)}$$

The Q_t, L_t, and I_v are observed data; while a, b, c, d, h, k, the m(t) and the $I_{m(t),t}$ are the unknowns to be estimated.[8] Production and employment data for the 1943–70 period are presented and discussed in Appendix Table 3. Investment data for 1919–69, together with an explanation of their construction, are available on request from the author.

A weighted first differencing of equation (6) yields:

$$(8) \qquad Q_{t+1} = e^{-d} Q_t + \frac{e^{kt}}{b} I_t + R_t$$

where

$$(9) \quad R_t = \frac{e^{km(t+1)}}{b} I_{m(t+1),t+1} - \frac{e^{-d}e^{km(t)}}{b} I_{m(t),t} + \begin{cases} -e^{-d} \displaystyle\sum_{v=m(t)+1}^{m(t+1)} \frac{e^{kv}}{b} e^{-d(t-1-v)} I_v & \text{if } m(t+1) > m(t) \\ 0 & \text{if } m(t+1) = m(t) \\ + \displaystyle\sum_{v=m(t+1)+1}^{m(t)} \frac{e^{kv}}{b} e^{-d(t-v)} I_v & \text{if } m(t+1) < m(t) \end{cases}$$

The first stage of estimation treats the remainders R_t as approximation errors and uses a non-linear least squares procedure to find that triplet $(\hat{b}, \hat{k}, \hat{d})$ which minimizes the sum of the squared errors:

$$\sum_t \left(Q_{t+1} - e^{-d} Q_t - \frac{e^{kt}}{b} I_t \right)^2$$

The remainder R_t may be viewed as the sum of a "true" approximation error R_t^* plus a stochastic error u_t, where R_t^* is defined from equation (9) under the true parameter values (b^*, k^*, d^*). If the total error terms $R_t^* + u_t$ are independent and

[8] The model can be made more general by replacing (5) with

$$(5') \qquad b_{v,t} = be^{-kv}e^{-ft}$$

where f represents the rate of disembodied decline in capital-output ratios over time. Note then that

$$b_{v,t} = be^{-f}e^{-(f+k)v}e^{-f(t-1-v)}$$

and equation (6) must be replaced by

$$(6') \qquad Q_t = \sum_{v=m(t)+1}^{t-1} \frac{e^{(f+k)v}}{be^{-f}} e^{-(d-f)(t-1-v)} I_v$$

(6′) continued

$$+ \frac{e^{(f+k)m(t)}}{be^{-f}} I_{m(t),t}$$

Thus, estimates of the triplet (b, k, d) under (5) would correspond to estimates of $(be^{-f}, f+k, d-f)$ under (5′).

normally distributed with zero means, then the triplet $(\hat{b}, \hat{k}, \hat{d})$ is the maximum likelihood estimate of (b^*, k^*, d^*).

The validity of assuming that the error terms have zero means is questionable. Under plausible values of (b, k, d), R_t tends to be negative whenever $m(t+1)>m(t)$, tends to be close to zero whenever $m(t+1)=m(t)$, and tends to be positive whenever $m(t+1)<m(t)$. But unless growth of output and investment allow a strong trend of increasing marginal vintage age, the case $m(t+1)>m(t)$ is expected to occur more frequently than the case $m(t+1)<m(t)$. For example, a symmetric distribution of $m(t+1)$ about $m(t)$ with $m(t+1)=m(t)+1$ for 5 out of 20 observation years, $m(t+1)=m(t)$ for 10 years, and $m(t+1)=m(t)-1$ for the remaining 5 years would increase the marginal vintage age by 20 years during the 20-year observation period. Obversely, the marginal vintage age will stay constant only if $m(t)+1$ is the observed mean value of $m(t+1)$. But if $m(t+1)=m(t)+1$ is indeed the expected case, R_t is more likely to be negative than positive. Multiple sampling of the Q_{t+1}, Q_t, and I_t for any particular t would not be likely to yield sample values of $R_t=R_t^*+u_t$ that were normally distributed about a zero mean.

This consideration suggests that the least squares estimates are not unbiased maximum likelihood estimates. To eliminate some of the bias, five cases in which $Q_{t+1}<.98Q_t$ were discarded from the sample; because a large decline in output is likely to induce a decline in marginal vintage age (i.e., $m(t+1)>m(t)+1$), and the associated approximation errors would then be particularly large and negative. Nevertheless, discarding these observations does not guarantee a small bias in the least squares estimates for the remaining observation years.[9]

The least squares estimates for the data period t=49–52, 54, 55, 58, and 60–66 are:[10]

$$(10)\quad (\hat{b}, \hat{k}, \hat{d}) = (.585, .00347, 5.96 \times 10^{-8})$$

While these values cannot be defended well on statistical grounds, on economic grounds they seem quite plausible. The estimate $\hat{d}\approx 0$ says there is virtually no physical depreciation of capital; but rather implies that retirement is attributable almost entirely to technological obsolescence.[11] The estimates $\hat{b}=.585$ and $\hat{k}=.00347$ imply $\hat{b}_v=.585e^{-.00347v}$; so $\hat{b}_{60}=.493$. This may be compared with other data, under the assumption that b_{60} approximates the average capital-output

[9] I am not aware of any formal statistical criterion for judging how seriously these approximation errors will bias the estimates of the non-linear model (8). Attiyeh and Barr have used different estimation procedures. Attiyeh was forced by data limitations to select as his estimates the solution to n equations in n unknowns, a procedure which involves no statistical degrees of freedom. Barr has made use of the assumption that industry average labor compensation represents a uniform competitive wage equal to value-added per worker on the oldest and marginally profitable vintage in operation. Attiyeh's method is clearly undesirable when data are adequate. Barr's method is theoretically elegant; but in fact there are sufficient regional differentials in textile earnings that such a procedure would probably understate marginal vintage age by at least 5 years. Barr's method is also difficult to apply in the open economy context. See Isard (pp. 88–91, 100) for further discussion.

[10] Further stages of estimation make the approximation $\hat{d}=0$.

[11] Throughout the data period, most of the capital estimated to have been in operation was younger than 20 years; and the estimate $\hat{d}\approx 0$ is probably very sensitive to this age structure. Since capital does not last forever, the estimate of negligible depreciation on young capital makes it appealing to use a "one-hoss shay" assumption in the projections of Section II, as will be discussed further below. It may also be noted that the oldest operating age of any vintage during the data period is estimated to have been 32 years (see Table 1 below). However, the shape of the age pyramid suggests that the empirical implications of the model are not very sensitive to the rate of depreciation on capital older than 20 years, and consequently would not be much affected if this rate of depreciation was distinguished and estimated separately from the rate of depreciation on capital younger than 20 years. Finally note that under the generalization outlined in fn. 8, this estimate must be reinterpreted as $\hat{d}-\hat{f}\approx 0$. Physical depreciation is then almost entirely offset by disembodied change in capital-output ratios.

ratio for 1960–61 (or that the weighted-average age of capital in use during 1960–61 was approximately 10 years). There should then be agreement between $\hat{b}_{60}$ and data from *U.S.* corporation income tax returns, which record that the undepreciated book value of textile capital was .415 times textile business receipts in 1960–61.[12] The discrepancy may result in large part from a failure to adjust book values upwards in response to the inflation that has occurred in capital goods prices.

Given the parameters (10), unique values of the m(t) and the $I_{m(t),t}$ can be computed from (6). Table 1 shows the implied ages and birthdates of the marginal vintages in operation during each observation year. The gradual increasing trend in marginal vintage ages can be attributed to an extension of profitable operating lives, due in part to the fact that unit labor requirements fell faster from vintage to vintage than real wages increased from year to year. In addition, rapid demand growth during the 1960's made it profitable for idle mills to resume operations, which also shows up in the marginal vintage age estimates.

As a further test of the estimates, the cyclical behavior of the implied marginal vintage ages may be compared with statistics on unfilled orders, also shown in Table 1. An increase in unfilled orders is expected to draw previously idle vintages back into operation (through cyclical influences on prices and profit rates), thereby increasing the marginal vintage age; and vice versa. This expectation seems confirmed. In only two out of nineteen cases do marginal vintage age and unfilled orders change in opposite directions.

The implied values of the m(t) and the $I_{m(t),t}$ can now be used along with (10) and (1) to compute first the implied $I_{v,t}$ and then that estimate of (a, c, h) which minimizes the sum of the squared errors in (7):

$$\sum_{t=48}^{67}\left(L_t - \frac{ae^{-ct}}{\hat{b}}\sum_{v=m(t)}^{t} e^{(\hat{k}-h)v}I_{v,t}\right)^2$$

The least squares estimate is:

$$(11)\quad (\hat{a}_0, \hat{c}_0, \hat{h}_0) = (1.463, .0403, .0024)$$

The values of $\hat{h}_0$ and $\hat{c}_0$ are rates of change per year; while $\hat{a}_0$ is measured in man-hours per constant (1957–59) dollar's worth of output.

The estimate (11) attributes most of the decline in labor-output ratios to disembodied efficiency increases ($\hat{c}_0 = .0403$), and suggests that very little of the observed change has been embodied in new capital vintages ($\hat{h}_0 = .0024$). This does not conform with a priori expectations. In the

TABLE 1—ESTIMATED BIRTHDATES AND AGES OF MARGINAL VINTAGES IN OPERATION, AND UNFILLED ORDERS AT THE END OF THE PREVIOUS YEAR, 1948–67

Year of operation t	Birth year of marginal vintage in operation m(t)	Age of marginal vintage in operation t−m(t)	Unfilled orders at the end of year t−1 (in weeks of equivalent production)[a]
48	24	24	16.2
49	26	23	5.8
50	26	24	12.6
51	27	24	15.6
52	28	24	9.4
53	29	24	10.1
54	32	22	7.1
55	30	25	10.2
56	31	25	13.6
57	34	23	10.5
58	36	22	11.1
59	32	27	13.5
60	35	25	21.5
61	36	25	9.9
62	35	27	11.9
63	36	27	11.1
64	36	28	12.8
65	34	31	18.2
66	34	32	20.3
67	38	29	18.2

[a] Office of Business Economics, *Business Statistics*, Washington 1967.

[12] See Internal Revenue Service.

opinion of Solow, purely disembodied change

> does not correspond very well with what actually goes on in the world under the name of technological progress. . . . it is not at all a metaphysical question whether technical progress is predominantly embodied or disembodied; whether gross investment is or is not the major vehicle by which new knowledge makes its way into the process of production. If it is not, then a stationary economy, or even one whose conventionally measured gross investment is zero, will find its productivity increasing as rapidly as a growing economy which allocates a substantial fraction of its capacity to conventionally-measured net investment. If it is, then acts of investment serve the double purpose of adding to the stock of capital goods available for future production and making that stock a little more modern, or raising the average level of technology available for use. [1963, pp. 42–44]

Moreover, such a low rate of embodied change ($\hat{h}_0 = .0024$) implies little difference between unit labor requirements on new and old capital vintages (for example, $\hat{a}_{t-1,t} = .93\ \hat{a}_{t-30,t}$). And with only slight embodied decline in capital-output ratios, it seems difficult to rationalize the observation of high rates of investment in an industry with idle plant and equipment.

For these reasons it was decided to examine how the statistical quality of the estimate would be affected by constraining technical change to be predominantly embodied. Two alternative estimates were computed, corresponding to the hypotheses: (i) that the change in labor-output ratios is purely embodied (i.e., $c=0$); and (ii) that the change in labor-output ratios from one year's newest vintage to the next year's newest vintage, $c+h$, is approximately two-thirds embodied (specifically, $h=.03$). These alternative hypotheses, respectively, yield the least squares estimates

$$(\hat{a}_1, \hat{c}_1, \hat{h}_1) = (.893, 0, .0436) \tag{12}$$

and

$$(\hat{a}_2, \hat{c}_2, \hat{h}_2) = (1.153, .0148, .03) \tag{13}$$

Neither of two statistical comparisons provides more than a very weak basis for choosing between the three estimates (11)–(13).[13]

Because of the difficulty in choosing a preferred estimate, the policy implications of each of the three estimates are presented below. It is nevertheless tempting to assume that technical change is at least two-thirds embodied, in which case the following computations suggest the desirability of policy revision. Under the hypothesis that technical change is approximately

[13] The first comparison computes for each estimate the square root of the mean squared deviation between observed and predicted employments (conventionally called the root mean squared error) as a fraction of the mean observed employment. At the estimates (11), (12), and (13), this statistic takes the respective values .0175, .0603, and .0403. By subtracting these values from unity, measures are obtained of the degree of variance explained by the three estimates. Thus, the estimate (11) explains 98.2 percent, (12) explains 94 percent, and (13) explains 96 percent.

The second comparison tests the accuracy of each estimate in predicting employment levels for 1968–70, the three years following the data period used in estimating the parameters. Each of the estimates consistently underpredicts employment, with predictions becoming progressively worse over time. In part, this may reflect a reluctance of employers to lay off workers during a downturn; and for 1970 it may reflect the additional difficulty of reducing work shifts below 40 hours per week. (Average weekly hours of production workers were 41.2, 40.8 and 39.9 during the three successive years. See U.S. Department of Labor.) On average, the estimate of purely embodied change yields the most accurate predictions, although it is judged the least accurate by the first test above.

	Predicted minus observed employment as a fraction of observed employment		
Year t	Based on (11)	Based on (12)	Based on (13)
68	−.063	−.026	−.059
69	−.074	−.037	−.071
70	−.091	−.100	−.117
Average	−.076	−.054	−.082

two-thirds embodied, the labor-output ratio on the most modern plant and equipment operated during 1970 is estimated to have been 49 percent of the unit labor requirement in the oldest and marginally profitable mills. Similarly, under the hypothesis of purely embodied change, modern mills in 1970 are estimated to have employed 35 percent as much labor per unit output as the oldest operating mills. Therefore, if technical change is at least two-thirds embodied, these estimates suggest that between 51 and 65 percent of the employment impact of 1970 imports could have been neutralized by cutting back 1969 investment in anticipation of the imports, thereby allowing imports to substitute for the additional output of modern mills instead of forcing marginally profitable mills to close. Obversely, such a cut back in new investment would have allowed 1970 imports to increase by an estimated factor between 2.04 $(=1/.49)$ and 2.86 $(=1/.35)$ without affecting 1970 employment.

II. Toward an Efficient Policy for the 1971–85 Period

Under a welfare criterion associated with Pareto, it is widely accepted that no policy should be undertaken unless it is "efficient" in the sense that no other policy could increase the well-being of any one group of economic participants without reducing the well-being of any other group. In practical judgement of economic policy, however, it is difficult to take into account the well-being of all economic participants, or even to define precisely the well-being of those participants who are most affected by the policy in question. Nevertheless, economists generally accept the somewhat vague proposition that it is efficient to expand domestic production of a tradeable good only if the value-added in production is sufficient to pay domestic factors of production their opportunity costs when the product (and each tradeable input) is valued at its world market price.[14]

According to this criterion, it does not seem efficient to encourage the expansion of the *U.S.* textile sector, which requires tariff and quota protection to attract its factors of production. On the other hand, to the extent that existing textile resources are specific to the industry, it is Pareto-efficient to contract domestic textile resources only if vested resources can be compensated for their losses. This suggests that an efficient *U.S.* textile policy is one which liberalizes imports in a manner controlled to phase out the domestic textile sector at the rate at which existing textile capital is scheduled for retirement,[15] and which is accompanied by demand-creation programs to insure a compensating number of new jobs in the regions of import-created unemployment.

Contrary to these considerations, *U.S.* policy encouraged the textile sector to undertake substantial new investment and additions to employment during the 1961–68 period. In 1961, the Kennedy Administration requested that *GATT* meet to seek an "international understanding" on textile trade,[16] which led to the Long Term Arrangement Regarding International Trade in Cotton Textiles (*LTA*).[17] Under the

[14] The "world market price" of a product is not, however, a well-defined concept.

[15] Or can be sold to other countries or converted to other uses without substantial loss to present owners.

[16] This was part of a seven point program of assistance to the *U.S.* textile industry, which represented both a fulfillment of campaign promises and part of a deal in which the textile industry agreed not to lobby against the Trade Expansion Act and the subsequent Kennedy Round negotiations for liberalizing international trade. See Thomas Curtis.

[17] The *LTA*, which became effective on October 1, 1962, after a one-year Short Term Arrangement, relates only to textile and apparel products in which cotton represents more than 50 percent of the fiber content, by weight. Under the terms of agreement, exports of such products may be limited only when they cause "disruptions" in markets of importing countries. According to *GATT*:

These situations [market disruptions] generally con-

LTA, the United States has imposed 5 percent quotas on the annual growth rate of cotton textile imports from most foreign suppliers outside the Western industrial bloc. Introduction of this import protection followed a liberalization of depreciation allowances on textile investments in late 1961, which reduced the average useful tax life of textile mill equipment from 25 to 14 years (and to 12 years for finishing equipment). These measures provided incentives for new textile investments, which were amplified by rapid 5 percent average annual growth of real domestic textile demand during the 1961–68 period. Investment skyrocketed! From annual rates in the range of $300 million constant (1958) dollars during the 1953–61 years, spending on new textile plant and equipment rose steadily to approximately $800 million in 1966, and averaged over $500 million during the 1967–69 years. By mid-1966 this was cause for alarm. To quote Curtis:

> The great increase of productive capacity for the U.S. textile industry is one of the key elements of concern about the future of that industry. The wiser policy would seem to be to unshackle import competition now, through more liberal administration of the long-term arrangement. . . . The real danger is that the capacity being created will only increase pressure to protect this new capacity later even though it be inefficient. It is no favor to the wage earner to entice him into a job which has an unstable economic base. [pp. 20988–89]

By 1968, *U.S.* textile employment had increased to 10.2 percent above its 1961 level, despite more than a 20 percent increase in the average real product of production workers. These new vested interests suffered from the decline in domestic textile demand between 1968 and 1970, which was accompanied by an acceleration of imports of man-made fiber products. As a consequence, the Nixon Administration forced new limitations on textile imports from Japan, Hong Kong, South Korea, and Taiwan.[18] Unless modified in the interest of an efficient long-run textile policy, these new limitations on woolens and man-made fiber products, together with continued use of cotton textile quotas under the *LTA*, are likely to foster a renewed expansion of the domestic textile sector once the *U.S.* economy again begins to exhibit a healthy real growth rate.

Suppose, however, that it is indeed possible to convince political pressure groups of the wisdom of an efficient textile policy. What would constitute an efficient policy for the 1971–85 period?

The answer to this question requires a judgement on the maximum rate at which the *U.S.* textile sector can be contracted without imposing "significant" or "politically unacceptable" costs on existing vested interest groups. This in turn de-

tain the following elements in combination:

(i) a sharp and substantial increase or potential increase of imports of particular products from particular sources;

(ii) these products are offered at prices which are substantially below those prevailing for similar goods of comparable quality in the market of the importing country;

(iii) there is serious damage to domestic producers or threat thereof; . . . .

In connection with sub-paragraph (iii) the "damage" referred to must be damage caused directly by market disruption and not by any change of consumer taste, technological advance, or similar factors. [pp. 14 and 22]

In practice, the United States in particular has interpreted any acceleration of imports as market disruption, even though imports through 1970 have never supplied more than 5 percent of domestic consumption. There have been loud protests from exporting countries, who have twice been forced into renewing the *LTA*, which now extends until October 1973. See Isard, ch. 3.

[18] These limitations cover eighteen categories of textiles and apparels consisting of man-made and wool fibers. Japan has agreed to limit the rate of growth of exports of these products to 5 percent per year, for 3 years; while the other three countries have agreed not to expand their exports by more than 7.5 percent per year, for 5 years. The agreements became effective as of October 1, 1971, with the base year for measurement the 12 months ending March 31, 1971. See *The New York Times*, Oct. 16, 1971, and *The Economist*, Oct. 23, 1971.

pends very much on the potency with which demand-creation programs can be designed to shift workers out of the textile sector into other activities.

For lack of such a judgement, the remainder of this section will consider several alternative rates of contraction of the size of the *U.S.* textile sector, as measured by employment. Under "pessimistic," "preferred," and "optimistic" forecasts of the rate of growth of textile demand, the vintage model and technical change estimates from Section I will be used to compute the growth of imports that is consistent with these alternative rates of employment decline, under an investment-retirement policy designed to hold at approximately 25 years the age of the oldest operating capital vintage. The choice of 25 years as a target retirement age is judged to be approximately optimal insofar as it utilizes existing capital for close to its "maximum practical lifespan" before replacing it with modern plant and equipment.[19]

[19] The estimate (10) suggests that there is almost no physical depreciation on young capital (recall fn. 11), which in turn suggests that once depreciation starts, it proceeds quite rapidly. This makes it tempting to think of capital as a one-hoss shay, which remains as good as new throughout its entire lifespan and then disintegrates instantaneously. If one were to vary the lifespan of the one-hoss shay until its depreciation curve best fit the "true" depreciation curve of textile capital, one would arrive at what I have chosen to call the "maximum practical lifespan" of textile capital. There are several foundations for the conjecture that 25 years approximates this maximum practical lifespan. First, even though Table 1 estimates that vintages as old as 32 years were in operation during the 1961–68 boom, when idle mills were reopened for production, the rise in the ratio of equipment to plant expenditures since World War II suggests a shorter practical lifespan for postwar investments. Second, independent evidence from corporate income tax records for 1954 and 1955, "the latest years in which reported depreciation allowances were largely free from the effects of the adoption of curvilinear depreciation methods under the Internal Revenue Code of 1954," show that depreciation charges by textile firms were consistent with the choice to use physical assets for an average life of 19 years, as opposed to 29 years during the prewar period (see Bert Hickman, pp. 238–43). This provides a lower bound on the maximum practical life.

The base year for projection is 1970, when domestic textile output (Q_{70}), domestic plus export demand (D_{70}), imports (M_{70}) and production-worker employment (L_{70}) were at the levels:

$$Q_{70} = 21.070 \text{ billion constant (1957–59) dollars}$$

$$D_{70} = 22.042 \text{ billion constant (1957–59) dollars}$$

$$M_{70} = 0.972 \text{ billion constant (1957–59) dollars}$$

$$L_{70} = 1.757 \text{ billion production-worker man-hours}$$

Twenty-five capital vintages are estimated to have been in operation to produce Q_{70}, so that average output per vintage was .84 billion. Since the maximum practical lifespan of textile capital has been conjectured to also be 25, it is appealing to consider the simple retirement policy which in each year closes down just enough of the oldest plant and equipment to reduce output capacity (net of new additions to capacity) by .84 billion.[20] When such a retirement policy is combined with new investment at the level I_t, the resulting change in employment between years t and t+1, under full capacity operation, can be approximated as:

$$(14) \quad \begin{aligned} L_{t+1} - L_t = &- .84ae^{-ct}e^{-h(t-25)} \\ &- (L_t - .84ae^{-ct}e^{-h(t-25)})(1 - e^{-c}) \\ &+ \frac{I_t}{be^{-kt}} ae^{-c(t+1)}e^{-ht} \end{aligned}$$

The top line on the right-hand side of this equation represents the employment loss on retired capacity, the second line is the employment loss due to disembodied tech-

[20] It can be computed that during the 1971–85 period such a simple retirement policy would result in the following frequency distribution of marginal vintage ages: 26 years, 3 times; 25 years, 7 times; 24 years, 3 times; 23 years, 2 times.

nical change on all but the most modern vintage, and the third line is the employment gain on new additions to capacity.

Suppose now that $U.S$ policy makers choose to phase out textile employment at the exponential rate z per year, or that the target employment path is

$$(15)\quad L_t = e^{-z(t-70)}L_{70} \quad \text{for } t = 71, \ldots, 85$$

Under the simple retirement policy proposed above, equation (14) defines a stream of investments that will keep textile employment approximately on this target path. In particular, by substituting (15) into (14) and then solving for I_t, this target investment stream is seen to be:

$$(16)\quad I_t = .84be^{25h}e^{-kt} + \frac{b}{a}(e^{c-z} - 1)e^{70z}e^{(c+h+k-z)t}L_{70} \quad \text{for } t = 70, \ldots 84$$

The problems of designing policies or coordination mechanisms to keep new investment (and retirement) on target might in practice prove to be among the more difficult and interesting aspects of an efficient textile policy; but here these problems are by-passed to focus instead upon the question of import liberalization.

If new investment and retirement are kept on target, full capacity output ($\overline{Q}$) of all operable capital follows the time path:

$$(17)\quad \overline{Q}_{t+1} = \overline{Q}_t - .84 + \frac{I_t}{be^{-kt}} \quad \text{for } t = 70, \ldots, 84$$

Adherence to the target employment path requires that domestic production be maintained at these full capacity levels. Thus, if D_{t+1} represents the sum of domestic textile demand plus the demand for $U.S.$ textile exports in year t+1, the level of imports (M_{t+1}) consistent with target employment is

$$M_{t+1} = D_{t+1} - \overline{Q}_{t+1}$$

Alternatively, the increase in imports that is consistent with adherence to target employment is

$$M_{t+1} - M_t = (D_{t+1} - D_t) - (\overline{Q}_{t+1} - \overline{Q}_t)$$

or from (17),

$$(18)\quad M_{t+1} - M_t = D_{t+1} - D_t + .84 - \frac{I_t}{be^{-kt}} \quad \text{for } t = 70, \ldots, 84$$

The growth of demand must be forecast to estimate in advance the increase in imports that is consistent with target employment.

In practice, policy could be designed to announce the rate of import liberalization on a short-term basis, thereby making use of short-term forecasts to predict cyclical movements in demand, and also allowing for adjustments to compensate for errors in previous forecasts of demand growth. Nevertheless, in order to estimate the long-run development assistance implications of any particular target rate of employment decline, and in order to allow exporting countries to forecast the long-run growth of the $U.S.$ market for their products, it is desirable to calculate the import growth that is consistent with long-run forecasts of the growth of demand.

The problems of forecasting the long-run growth of demand have been dealt with in my doctoral dissertation. Exports constitute less than 5 percent of the combined domestic plus export demand, so that the growth rate of domestic demand is a good estimate of the growth rate of the combined demand. Because of the historically erratic nature of domestic demand growth, three alternative forecasts are adopted here. A pessimistic forecast predicts a 2 percent annual exponential rate of growth; a preferred forecast predicts a 3.5 percent rate; and an optimistic forecast predicts 5 percent. The pessimistic rate corresponds roughly to the experience of

the 1950's; and the optimistic rate reflects the 1961–68 experience. The preferred 3.5 percent rate corresponds to a per capita income elasticity of .80 to .83 if real *GNP* grows at 4 percent per year, or .66 to .71 if real *GNP* grows at 4.5 percent, using the range of population projections that are published by the Census Bureau.[21] There is some evidence that *U.S.* industry is basing its future plans on predictions which lie between the preferred and optimistic forecasts.[22]

Under the annual exponential growth rate g, domestic plus export demand in year t is:

$$D_t = e^{g(t-70)} D_{70} \tag{19}$$

and equation (18) becomes

$$M_{t+1} - M_t = (e^g - 1)e^{g(t-70)} D_{70} + .84 - \frac{I_t}{be^{-kt}} \quad \text{for } t = 70, \ldots, 84 \tag{20}$$

Table 2 shows the total growth in imports over the 1971–85 period (i.e., the value of $M_{85}-M_{70}$) that is consistent with target employment under five alternative target rates of employment decline ($z=0$, .015, .03, .045, .06), three alternative forecasts of the rate of demand growth ($g=.02$, .035, .05), the assumption that technical change in capital-output ratios will procede according to the estimate (10), and three alternative hypotheses regarding technical change in labor-output ratios (the parameters of (11)–(13)).

The table suggests that if long-run demand growth slackens from the 5 percent annual pace witnessed during the 1961–68 period, import liberalization is not consistent with a constant (or increasing) level of employment. However, if a target rate of employment decline of at least 3 percent per year is chosen, then the preferred forecast of 3.5 percent annual demand growth allows for substantial growth of imports under any of the three hypotheses regarding technical change in unit labor requirements. It is worth stressing the fact that the import implications are fairly insensitive to the nature of technical change when employment is reduced by at least 3 percent per year. Under the 3 percent target, the preferred demand forecast suggests that imports can increase to more than 11 times their base year level ($M_{70}=.97$) in 1985. Each year would allow imports to increase by an average of at least .67 billion over the previous year's level; and on the average, each year's import total would be at least 5 billion dollars greater than the level of 1970 imports. With preferential trade agreements, if necessary, to insure that a substantial fraction of the additional import orders were captured by the *LDC*s, such a policy would provide a significant amount of development assistance.

III. Qualifications and Conclusions

Several crude assumptions have been adopted in structuring and estimating the vintage model of Section I: the available

[21] If N represents population and g_{DD}, g_N, and g_{GNP} respectively, represent the growth rates of domestic demand, population, and *GNP*, then the per capita income elasticity a is defined implicitly by

$$\frac{DD}{N} = c\left(\frac{GNP}{N}\right)^a$$

where c is a constant; and a is estimated from the implied relationship

$$a = \frac{g_{DD} - g_N}{g_{GNP} - g_N}$$

The Bureau of the Census (1971) forecasts population to grow at an average annual exponential rate between 1.05 percent (the Series E projection) and 1.54 percent (Series B) during the 1970–85 period.

[22] Charles B. McCoy, President of Dupont, the largest man-made fiber producer, has predicted that *U.S.* consumers will increase their purchases of textile products by at least 50 percent between 1970 and 1980 (see *The New York Times*, March 21, 1970). The preferred forecast predicts that demand will increase by 42 percent every 10 years, while the optimistic forecast predicts a 65 percent increase.

TABLE 2—THE IMPORT IMPLICATIONS OF ALTERNATIVE TARGET EMPLOYMENT PATHS UNDER APPROXIMATELY EFFICIENT INVESTMENT AND RETIREMENT POLICIES

Target rate of employment decline (z)	Forecast growth rate of demand (g)	Hypothesized form of technical change: Disembodied, see equation (11)	Embodied, see equation (12)	Two-thirds embodied, see equation (13)
		Associated Levels of $M_{85}-M_{70}$[a]		
	.02	−13.3	−17.2	−17.3
0.0	.035	− 5.7	− 9.7	− 9.8
	.05	3.7	− 0.3	− 0.4
	.02	− 4.3	− 5.7	− 6.2
.015	.035	3.2	1.8	1.3
	.05	12.6	11.2	10.7
	.02	2.9	3.2	2.5
.03	.035	10.4	10.7	10.0
	.05	19.8	20.1	19.4
	.02	8.6	10.0	9.3
.045	.035	16.1	17.6	16.8
	.05	25.5	27.0	26.2
	.02	13.1	15.3	14.6
.06	.035	20.6	22.8	22.1
	.05	30.0	32.2	31.5

[a] In billions of constant (1957–59) dollars.

data did not enable a disaggregation of plant and equipment investment; and the most acceptable method of estimating the model involved the strong error assumption adopted in fitting equation (8). Moreover, the policy alternatives investigated in Section II are based upon a fairly crude conjecture regarding the maximum practical use that can be obtained from textile plant and equipment. Thus, Table 2 is offered as a rough approximation of the maximum import growth that is feasible under alternative target employment paths, for a range of demand growth and technical change contingencies.

Despite the room for empirical refinement, the tenor of the conclusions is strong. With 1970 textile imports representing less than 5 percent of the *U.S.* market, there is substantial room for development assistance through import liberalization. Such development assistance can be made available to *LDC*s through a restructuring of policy toward the long-run goal of gradually phasing out the domestic textile sector. If new investment and capital retirement are controlled in a manner designed to obtain (close to) the maximum practical use from existing plant and equipment, the import levels consistent with any target path of employment decline will be (approximately) maximized. In particular, on the assumption that technical change in labor-output ratios is at least two-thirds embodied, the import levels that are consistent with any decline in textile employment can be increased by an estimated factor of 2.04 to 2.86 if new investment is cut back, in anticipation of the imports, in order to avoid the closure of marginally profitable mills with relatively high unit labor requirements.

Fortunately, policy formulation does not require an extremely high degree of confidence in the estimated economic implications of each alternative. Policy can be adjusted over time as new information becomes available to challenge the wisdom of the target retirement age, the predictions of demand growth, or the accuracy of the estimated capital-output and labor-output ratios. In practice, the major difficulty of designing a policy to phase out comparatively disadvantageous production activities is the strong opposition from groups with vested interests in these activities. The challenge is to construct complementary policies to compensate for the

Data Appendix

Table 3—*U.S.* Textile Employment, Production, Imports, Exports, and Domestic Demand, 1948–70

Year	Employment[a]	Production[b]	Imports[c]	Exports[c]	Domestic Demand[d]
1948	2.544	12.857			
1949	2.157	11.894			
1950	2.407	13.537			
1951	2.313	13.395			
1952	2.182	13.282			
1953	2.163	13.551			
1954	1.898	12.673			
1955	2.005	14.004			
1956	1.949	14.217	.409	.442	14.184
1957	1.807	13.664	.386	.461	13.589
1958	1.671	13.353	.388	.435	13.307
1959	1.801	15.463	.491	.449	15.504
1960	1.715	14.868	.556	.473	14.951
1961	1.670	15.151	.505	.470	15.187
1962	1.715	16.327	.644	.487	16.484
1963	1.675	16.553	.669	.484	16.738
1964	1.702	17.403	.670	.570	17.502
1965	1.797	19.102	.774	.514	19.362
1966	1.871	20.178	.866	.532	20.511
1967	1.805	20.107	.759	.496	20.370
1968	1.887	21.382	.867	.470	21.779
1969	1.869	21.835	.888	.501	22.221
1970	1.757	21.070	.972	.516	21.526

[a] Billions of production-worker man-hours per year. Constructed using data on production-worker employment and average weekly hours from U.S. Department of Labor.

[b] Billions of constant (1957–59) dollars. Constructed by multiplying the *FRB* production index for textiles (from Board of Governors of the Federal Reserve System) by .1416. This conversion factor was derived using data on value of shipments and end-of-year inventories from Office of Business Economics, *Surv. Curr. Bus.*, various issues, along with a textile consumer price deflator that was provided by the Bureau of Labor Statistics (*BLS*). The deflated sum of value of shipments plus increase in end-of-year inventories (in billions of 1957–59 dollars) was .1416 times as large as the *FRB* index in both 1963 and 1964, and .1417 times as large in 1965.

[c] Billions of constant (1957–59) dollars. Current dollar data are from: Bureau of the Census, *Statistical Abstracts*, various issues; Office of Business Economics, *Surv. Curr. Bus.*, various issues; and U.S. Tariff Commission. The deflator is the *BLS* consumer price index for textiles.

[d] Billions of constant (1957–59) dollars. Domestic demand equals domestic production plus imports minus exports.

losses of these groups, and thus to end their opposition. Without efforts to meet the challenge of restructuring production activities to conform with international comparative advantages, it is pure hypocrisy for politicians in the industrialized countries to express their interest in assisting development by promoting the export of *LDC* manufactures.

REFERENCES

R. Attiyeh, "Estimation of a Fixed Coefficient Vintage Model of Production," *Yale Econ. Essays*, spring 1967, *7*, 255–64.

J. L. Barr, "Empirical Implications of a Vintage Production Model," unpublished doctoral dissertation, Stanford Univ. 1969.

L. J. Chawner, "Capital Expenditures in Selected Manufacturing Industries," *Survey Curr. Bus.*, Dec. 1941, *21*, 19–26.

T. Curtis, "Toward a Modern U.S. Textile Import Policy," in U.S. Congress, *Congressional Record—House*, Washington, Aug. 29, 1966, 20966–96.

B. G. Hickman, *Investment Demand and U.S. Economic Growth*, Washington 1965.

G. C. Hufbauer, "Factor Endowments, National Size, and Changing Technology: Their Impact on the Commodity Composition of Trade in Manufactured Goods," in R. Vernon, ed., *The Technology Factor in International Trade*, New York 1970.

P. Isard, "The Development Assistance Content and Domestic Employment Implications of U.S. Textile Imports," unpublished doctoral dissertation, Stanford Univ. 1971.

L. Johansen, "Substitution Versus Fixed Production Coefficients in the Theory of Economic Growth: A Synthesis," *Econometrica*, Apr. 1959, *27*, 157–76.

D. B. Keesing, "Outward-Looking Policies and Economic Development," *Econ. J.*, June 1967, *77*, 303–20.

H. B. Lary, *Imports of Manufactures from Less Developed Countries*, New York 1968.

R. M. Nixon, "United States Foreign Policy for the 1970's: A New Strategy for Peace," report to Congress, published in *The New York Times*, Feb. 19, 1970.

R. M. Solow, "Substitution and Fixed Proportions in the Theory of Capital," *Rev. Econ. Stud.*, June 1962, *29*, 207–18.

———, *Capital Theory and the Rate of Return*, Amsterdam 1963.

———, J. Tobin, C. C. von Weizsäcker, and M. E. Yaari, "Neoclassical Growth with Fixed Factor Proportions," *Rev. Econ. Stud.*, Apr. 1966, *33*, 79–115.

Board of Governors of the Federal Reserve System, *Federal Reserve Bull.*, Washington, various issues.

The Economist, Oct. 23, 1971.

General Agreement on Tariffs and Trade (*GATT*), *Long Term Arrangement Regarding International Trade in Cotton Textiles*, Geneva 1963.

Internal Revenue Service, *Statistics of Income, 1960–61: Corporation Income Tax Returns*, Washington 1961.

The New York Times, Mar. 21, 1970 and Oct 16, 1971.

U.S. Department of Commerce, Bureau of the Census, *Census of Manufactures, 1963*, Vol. *2*, Part 1, Washington 1964.

———, *Annual Survey of Manufactures, 1966: Expenditures for New Plant and New Equipment*, Washington 1967.

———, *Current Population Reports: Population Estimates and Projections*, Ser. P-25, No. 470, Washington, Nov. 1971.

———, *Statistical Abstract of the United States*, Washington, various issues.

U.S. Department of Commerce, Office of Business Economics, *Business Statistics*, Washington 1967.

———, *1967 Capital Stock Survey*, Washington 1968.

———, *Survey of Current Business*, Washington, various issues.

———, *The National Income and Product Accounts of the United States, 1929–1965*, Washington 1966.

U.S. Department of Labor, *Manpower Report of the President*, Washington 1971.

U.S. Tariff Commission, *Textiles and Apparel*, TC Publication 226, Vols. *1* and *2*, Washington, Jan. 1968.

A Generalization of the Pure Theory of Public Goods

By Bryan Ellickson*

The concept of a pure public good formalized by Paul Samuelson (1954) has frequently been criticized for failing to admit the possibility of exclusion or crowding. In most public consumption—viewing a movie, belonging to a neighborhood, sending children to school, or sharing a dwelling—one shares consumption with some but not all other members of society. Adding another person to a sharing group while maintaining the quality of the service consumed will in general require additional resources. Most goods are public, but few are purely public.

This paper extends the concept of a public good to cover cases in which exclusion or crowding occurs. Like Samuelson, I assume that a public good is equally available to every member of a group to which it is provided. However, rather than assuming that everyone shares the public good, I permit the partition of the members of the economy into distinct sharing groups or "jurisdictions." The resources required to provide a public good to the members of a jurisdiction may depend on the size of the jurisdiction. Thus, both exclusion and crowding of public goods are admitted into the analysis.

The theory developed here not only permits the analysis of a wide class of public goods but sheds new light on certain results which have been obtained for the case of pure public goods. In particular, when crowding and exclusion are present, the Lindahl equilibrium need not be in the core and the core itself may be empty. These results stem directly from the fact that, within the context of this generalized model of public goods, the aggregate technology set may not be convex.

In contrast to the usual conclusion in economics, the presence of nonconvexity does not prevent further analysis of the allocative process in such an economy. By admitting nonconvexity, the theory is able to accommodate "neighborhood" and "local political jurisdiction" as legitimate analytical constructs. For a fixed assignment of individuals to jurisdictions, it is possible to prove the existence and relative Pareto optimality of a Lindahl equilibrium or of what Duncan Foley (1967) has called a public competitive equilibrium. However, because of the presence of nonconvexity, prices can guide the *assignment* of individuals to jurisdictions only under highly specialized circumstances.

Section I reviews the theory of pure public goods in terms of a three-person economy. Section II introduces crowding of the public good and the possibility of exclusion into this economy and exhibits cases where the Lindahl allocation lies outside of the core and where the core is empty, respectively. Section III explores the nature of the nonconvexity introduced by crowding and exclusion while Section IV interprets this nonconvexity as a type of indivisibility. After a brief discussion in Section V of the type of exclusion assumed in this analysis, Section VI turns to the general question of the relationship be-

* Assistant professor, University of California, Los Angeles. I have benefited greatly from comments by Joseph Newhouse, Joseph Ostroy, and a referee.

tween competitive analysis and the organization of collective consumption—the pattern of sharing public goods. Section VII demonstrates the failure of Lindahl prices to guide the assignment of individuals to jurisdictions. Section VIII relates the theory of public goods developed in this paper to James Buchanan's theory of clubs.

I. Pure Public Goods

Consider a three-person economy with two commodities, one private and one purely public. The tastes of the ith individual are represented by

$$(1) \qquad u_i = x_i g \qquad (i = 1, 2, 3)$$

where x_i is the amount of the private good and g the amount of the public good consumed. The ith individual has an initial endowment, w_i, of the private good; no public goods are initially owned; and public goods cannot be used as inputs to production. The economy faces the resource constraint

$$(2) \qquad cg + \sum_{i=1}^{3} x_i = \sum_{i=1}^{3} w_i$$

where c is a positive constant. The Pareto optimal level of the public good is then[1]

$$(3) \qquad g = \frac{w_1 + w_2 + w_3}{2c}$$

In their search for an analogue to competitive equilibrium in an economy with public goods, economists have given considerable attention to Lindahl equilibria (see Erik Lindahl, Leif Johansen, and Foley (1970)). The distinguishing feature of a Lindahl equilibrium is that taxes are levied in such a manner that each individual's "demand" for the public good will be brought into exact equality with the amount supplied. In our three-person economy, the level of public good supplied in the Lindahl equilibrium is given by equation (3). Setting the price of the private good arbitrarily equal to unity, the budget line of the ith individual will be tangent to his indifference curve at the Lindahl allocation provided that he is taxed an amount[2]

$$(4) \qquad t_i = \frac{w_i}{2} \qquad (i = 1, 2, 3)$$

Foley (1970) has proved under relatively general conditions, satisfied by our three-person economy, that the Lindahl equilibrium exists, that it is Pareto optimal, and that it belongs to the core. Foley puts great stress on the third result: no coalition can, using only its own resources achieve an allocation which makes each of its members better off than in the Lindahl equilibrium. Samuelson (1969) has voiced considerable skepticism concerning the relevance of Lindahl equilibria to the political-economic decision-making process. But as Foley observes: "There is some reason to think that the core is a meaningful political concept" (1970, p. 71).

II. Crowded Public Goods, Lindahl Equilibria and the Core

The case for the relevance of Lindahl equilibria would be stronger if one could prove some analogue of the Scarf-Debreu theorem on the core of a competitive economy. The Lindahl equilibrium would, in the limit, be the only politically relevant allocation. However, a theorem of this type does not seem possible when pure

[1] Equating the sum of marginal rates of substituion to marginal cost, $(x_1+x_2+x_3)/g=c$. Using equation (2), this relation can be translated into equation (3).

[2] To derive the Lindahl taxes, maximize $u_i=x_ig$ subject to the budget constraint $p_g^i g+p_x x_i=p_x w_i$ where p_g^i is the Lindahl price charged to the ith individual and p_x is the price of the private good. From the necessary conditions for a maximum, we obtain $p_g^i g=p_x x_i$. Substituting into the budget constraint yields $x_i=w_i/2$ and $p_g^i g=(p_x w_i)/2$. Letting $p_x\equiv 1$ and $t_i\equiv p_g^i g$ gives the desired result.

public goods are present. If the resource cost of producing a given level of the public good is independent of the number of individuals who consume it, then blocking is very difficult. In looking for ways to enhance the blocking ability of a coalition, it is natural to consider the possibility that the coalition can, by excluding nonmembers, obtain a given level of the public good at lower cost. When the addition of another consumer increases the resources required to maintain the level of public good consumed by all, we will refer to the public good as "crowded."

We will now introduce crowding into our three-person economy. If all three individuals consume the public good jointly, then we assume, as before, that they face the resource constraint given by equation (2). However, an individual who decides to produce the good g using only his own resources faces the constraint

$$ag_i + x_i = w_i \qquad (i = 1, 2, 3) \tag{5}$$

where a is a positive constant.[3] Any pair of individuals i and j $(i \neq j)$ face the constraint

$$bg_{ij} + x_i + x_j = w_i + w_j \tag{6}$$

provided that they share consumption of the public good.[4]

To be in the core, a vector of utilities (u_1, u_2, u_3) must be *individually rational*; i.e., every individual must obtain at least as much utility as he could using only his own resources. In our three-person economy, (u_1, u_2, u_3) is individually rational if, and only if,[5]

$$u_1 \geqq \frac{(w_1)^2}{4a} \equiv v(\bar{1}); \quad u_2 \geqq \frac{(w_2)^2}{4a} \equiv v(\bar{2}); \quad u_3 \geqq \frac{(w_3)^2}{4a} \equiv v(\bar{3}) \tag{7}$$

Although a pair of individuals forming a blocking coalition may choose to consume the good g privately rather than sharing it, we will, in order to simplify the exposition, assume that g is a pure public good for coalitions of size less than or equal to two; i.e., $a=b$. In that case, the set of utilities that a pair of individuals can obtain through private consumption of the "public" good is a proper subset of the set of utilities obtainable through joint consumption. Under these circumstances, equation (6) is the appropriate resource constraint for a two-person blocking coalition.

If (u_1, u_2, u_3) is to be in the core, then it must not be possible for any two-person coalition, using only its own resources, to obtain higher utility levels for both of its members. Therefore, (u_1, u_2, u_3) is in the core only if[6]

$$\begin{aligned} u_1 + u_2 &\geqq \frac{(w_1 + w_2)^2}{4b} \equiv v(\overline{12}) \\ u_1 + u_3 &\geqq \frac{(w_1 + w_3)^2}{4b} \equiv v(\overline{13}) \\ u_2 + u_3 &\geqq \frac{(w_2 + w_3)^2}{4b} \equiv v(\overline{23}) \end{aligned} \tag{8}$$

Returning to the case where all three individuals consume the public good jointly,[7]

[3] If the public good is uncrowded (i.e., if it is a pure public good) then $a=c$.

[4] If they consume the good g privately, then the resource constraint would be $a(g_i+g_j)+x_i+x_j=w_i+w_j$. To avoid an unnecessary complication, we will rule this possibility out by assumption.

[5] These expressions, as well as those of (8) and (9) below, are derived by determining the Pareto optimal "utility frontier" for each coalition using its resources alone and the resource constraint appropriate to its size.

[6] See fn. 5. In this case, the utility frontiers happen to be linear, but that will not be true in general even for three-person two commodity economies of the sort we are discussing. The counter-examples of this section can be obtained in the more general case, but the exposition would be unnecessarily complicated.

[7] We should consider other possible "partitions": e.g., 1 and 2 consuming jointly and 3 privately. But we rule out these other partitions by assuming that all individually rational utility vectors achievable under such a partition form a proper subset of the utility vectors determined by $u_1+u_2+u_3 \leqq v(\overline{123})$. One can show that

Pareto optimal utility vectors satisfy the equation

$$(9)\quad u_1 + u_2 + u_3 = \frac{(w_1 + w_2 + w_3)^2}{4c} \equiv v(\overline{123})$$

The inequalities (7) and (8) and equation (9) describe a type of game that is particularly easy to analyze: a game with transferable utility.[8] For any coalition S, the number $v(S)$ represents the minimum utility the coalition must receive if it is not to block. The number $v(S)$, defined over all possible coalitions, is called the characteristic function of the game.[9]

This game will have a nonempty core if, and only if,[10]

$$(10)\quad v(\overline{12}) + v(\overline{13}) + v(\overline{23}) \leqq 2v(\overline{123})$$

If (10) is satisfied, the game is said to be *balanced*.[11] Substituting (8) and (9) into (10), we obtain

$$(11)\qquad \frac{c}{b} \leqq \frac{k_1 + 2k_2}{k_1 + k_2}$$

as the necessary and sufficient condition for a nonempty core where $k_1 = (w_1)^2 + (w_2)^2 + (w_3)^2$ and $k_2 = w_1w_2 + w_1w_3 + w_2w_3$.

Having completed these preliminaries, we can now demonstrate the two propositions mentioned in the introduction: the Lindahl equilibrium is not necessarily in the core, and the core may be empty.

To establish the first proposition, let $a = b = 1$, $c = 4/3$, $w_1 = w_2 = 2$ and $w_3 = 1$. Substituting these values into (7), (8), and (9) we obtain the following game:

$$(12)\quad v(\overline{1}) = 1; \quad v(\overline{2}) = 1; \quad v(\overline{3}) = 1/4$$
$$v(\overline{12}) = 4; \quad v(\overline{13}) = 9/4; \quad v(\overline{23}) = 9/4$$
$$v(\overline{123}) = 75/16$$

From (11), the game will be balanced if $c \leqq 25/17$; since $c = 4/3$, the game has a nonempty core. A utility vector in the core must satisfy the inequality $u_1 + u_2 \geqq v(\overline{12}) = 4$, so the third individual can receive at most $v(\overline{123}) - 4 = 11/16$; individual rationality requires that he receive at least $v(\overline{3}) = 1/4$. Thus, for all utility vectors in the core, $1/4 \leqq u_3 \leqq 11/16$. (2, 2, 11/16) and (2, 39/16, 1/4) are, for example, two utility vectors that are in the core.

We have already observed that, in the Lindahl equilibrium,[12] $g = (w_1 + w_2 + w_3)/2c$ and $x_i = w_i/2$; consequently,

$$(13)\quad \bar{u}_i = \frac{w_i(w_1 + w_2 + w_3)}{4c} \qquad (i = 1, 2, 3)$$

For the particular parameter values we have selected, the Lindahl utility vector is $(\bar{u}_1, \bar{u}_2, \bar{u}_3) = (15/8, 15/8, 15/16)$. Since $\bar{u}_1 + \bar{u}_2 + \bar{u}_3 = 75/16$, the Lindahl equilibrium is a Pareto optimum. But individuals 1 and 2, who can obtain $(u_1, u_2) = (2, 2)$ using only their own resources, will block this Lindahl allocation. The Lindahl equilibrium fails to give sufficient recognition to the "power" of individuals 1 and 2—they receive too little and individual 3 receives too much—and, hence, the Lindahl allocation is not in the core. However, individual 3 is willing to allow individuals 1 and 2 to receive more than their "Lindahl share" if they share consumption of the public good with him: there are allocations involving joint consumption by all three individuals that are in the core.[13]

this condition will be satisfied if, and only if, the game as represented by (7), (8), and (9) is superadditive. In all of the examples to be presented, this superadditivity condition will be satisfied.

[8] As noted previously, utility is transferable only because the utility frontiers happen to be linear in the particular example.

[9] $v(S)$ is also the best that S can do assuming that $\overline{S}$ does its worst—the usual definition of the characteristic function.

[10] See L. S. Shapley. Strictly speaking, this condition will be sufficient only if the game is superadditive. This additional condition is satisfied in all of the examples to be discussed.

[11] For a discussion of balance in an economic context, see Herbert Scarf.

[12] See equation (3) and fn. 2.

[13] In fact, all allocations in the core involve sharing among all three individuals in this example.

We summarize this result in the following proposition:

PROPOSITION 1: *If public goods are crowded, then the Lindahl equilibrium may not be in the core.*

The second proposition that we wish to establish is that the core can be empty if public goods are crowded.[14] Letting $a=b=1$ and $w_1=w_2=w_3$ the balance condition (11) states that the core will be empty if $c>3/2$. If $c=3/2$ is interpreted as "constant returns to scale" in moving from two- to three-person sharing groups, then we find that the core is empty when there are diminishing returns to scale. One should not conclude, however, that the core will necessarily be nonempty whenever returns to scale are constant or increasing. For example, when $w_1=w_2=4$, $w_3=3$, $a=b=1$, and $c=3/2$, the core is empty.[15] In summary, we have demonstrated:

PROPOSITION 2: *If public goods are crowded, the core may be empty.*

III. Nonconvexity

To what can these rather surprising results be attributed? They stem from an assumption, implicit in the preceding section, that the aggregate technology set is nonconvex. Making this assumption explicit will be greatly facilitated if we formalize our approach to crowded public goods along the lines of Foley (1970).

We assume that the economy has one public good, m private goods, and n consumers. An *allocation* is a vector $(g_1, \ldots, g_n; x_1, \ldots, x_n)$ with each x_i itself a vector in E^m, such that for all i there is a $(g_i; \bar{x}_i)$ with $\bar{x}_i<x_i$ belonging to the consumption set of the ith consumer.[16] An allocation is *feasible* if $(g_1, \ldots, g_n; z)\in Y$ where Y is the aggregate technology set, w_i is the ith individual's initial endowment of private goods,[17] and $z=\sum_{i=1}^{n}(x_i-w_i)$ is the aggregate input of private goods into public good production. Of the assumptions which Foley makes about the aggregate technology set, the one which we will find inappropriate in an economy with crowded public goods is that Y is a convex cone.[18]

DEFINITION: A *Lindahl equilibrium* with respect to $w=(w_1, \ldots, w_n)$ is a feasible allocation $(g_1, \ldots, g_n; x_1, \ldots, x_n)$ and a price system $(p_g^1, \ldots, p_g^n; p_x)\geq 0$ such that

$$\text{(a)}\quad \sum_{i=1}^{n} p_g^i g_i + p_x\cdot z \geqq \sum_{i=1}^{n} p_g^i \bar{g}_i + p_x\cdot\bar{z} \quad \text{for all } (\bar{g}_1, \ldots, \bar{g}_n; \bar{z}) \in Y;$$

$$\text{(b)}\quad \text{if } (\bar{g}_i; \bar{x}_i)\succ_i (g_i; x_i) \text{ then } p_g^i\bar{g}_i + p_x\cdot\bar{x}_i > p_g^i g_i + p_x\cdot x_i = p_x\cdot w_i$$

Observe that there is a separate public good price, p_g^i, for each individual. In the

[14] This proposition runs counter to a theorem of Shapley and Martin Shubik that when external economies are present but external diseconomies are not then the core is always nonempty provided that utility and production functions satisfy the sufficient conditions for the existence of a competitive equilibrium. However, as we will see in the following section, the *aggregate* technology set will be nonconvex in this example. Results similar to Proposition 2 have been obtained by Shubik and by Joseph Ostroy.

[15] The reader can verify that this game is superadditive so that (9) is the appropriate characteristic function for $v(\overline{123})$.

[16] The assumptions on the individual consumption sets and preference orderings introduced by Foley, which are quite standard, will be adopted in this study. Each consumer chooses a point $(g_i; x_i)$ in a consumption set X_i on which there is defined a complete and transitive ordering $>_i$. The individual consumption sets are closed and convex and have an interior in the private good subspace. If $(g_i; x_i)\in X_i$, then there is a point $(g_i; \hat{x}_i)\in X_i$ with $\hat{x}_i<w_i$. The aggregate consumption set, $X=\Sigma_{i=1}^{n} X_i$, has a lower bound for $\leq$. The individual preference orderings satisfy assumptions of continuity, convexity, and monotonicity.

[17] No public goods are initially owned.

[18] We will adopt the other assumptions on Y introduced by Foley: Y is closed, $0\in Y$ (inaction is possible), no production other than inaction is possible without input of some private good, it is possible to produce the public good and the public good is unnecessary as a productive input.

Lindahl equilibrium, producers maximize profits and consumers maximize utility taking prices as given. Foley's definition of a Lindahl equilibrium is restricted to the special case of a pure public good where $g_1 = \ldots = g_n$. However, the concept of a Lindahl equilibrium, as defined above, can be applied to any pattern of sharing.

At this point, it is useful to formalize the notion of alternative patterns of sharing of the public good. Let $N \equiv \{1, \ldots, n\}$ denote the index set of consumers in the economy. Any subset of N representing a group of consumers who share consumption of the public good will be called a *jurisdiction*; the jth jurisdiction will be denoted J_j. Assume that every consumer belongs to exactly one jurisdiction. A particular assignment of consumers to jurisdictions, $\{J_1, \ldots, J_r\} \equiv P_k$, is a *partition* if

$$\bigcup_{j=1}^{r} J_j = N$$

The collection of possible partitions will be denoted by $\mathcal{P} = \{P_1, \ldots, P_s\}$. We reserve P_1 for the *private good partition* $(P_1 = \{\{1\}, \ldots, \{n\}\})$ and P_s for the *pure public good partition* $(P_s = \{N\})$. To any partition P_k corresponds an aggregate technology set $Y(P_k)$. In Foley's model of pure public goods, $Y = Y(P_s)$; in the standard model of private goods, $Y = Y(P_1)$.

Consider the three-person economy described in Section II. We assumed that the ith individual using only his own resources to produce the good g faced the resource constraint $ag_i + x_i = w_i (i = 1, 2, 3)$. Therefore, the production sets for the three consumers are the halflines[19] respectively associated with the vectors

$$\overset{1}{y}_1 = (1, 0, 0; -a) \tag{14}$$

$$\overset{1}{y}_2 = (0, 1, 0; -a)$$

$$\overset{1}{y}_3 = (0, 0, 1; -a)$$

where the vectors are written in the form $(g_1, g_2, g_3; z)$ with z representing the input, $x_i - w_i$, of private goods into the production of the public good. We also assumed that a pair of individuals $\overline{ij}$, using their own resources and sharing consumption of the public good, faced the constraint $bg_{ij} + x_i + x_j = w_i + w_j$. The production sets for the pairs $\overline{12}$, $\overline{13}$, and $\overline{23}$ are the halflines respectively associated with the vectors

$$\overset{2}{y}_1 = (1, 1, 0; -b) \tag{15}$$

$$\overset{2}{y}_2 = (1, 0, 1; -b)$$

$$\overset{2}{y}_3 = (0, 1, 1; -b)$$

Finally, we assumed that the constraint facing the three consumers if they consume the public good jointly is $cg + x_1 + x_2 + x_3 = w_1 + w_2 + w_3$; the production set is then the halfline associated with the vector

$$\overset{3}{y}_1 = (1, 1, 1; -c) \tag{16}$$

If we assume that, for a given partition P_k, the economy as a whole has the same technological capacity for producing goods for a component jurisdiction that the jurisdiction has when operating in isolation, then the aggregate technology set $Y(P_k)$ for the partition P_k is a convex cone.[20] In Section II, the partition P_k was

[19] The halfline (b) associated with the vector b is the set $\{x \mid x = \lambda b, \lambda \geqq 0\}$.

[20] Specifically,

$$Y(P_1) = \{y_i \mid y_i = \alpha \overset{1}{y}_1 + \beta \overset{1}{y}_2 + \gamma \overset{1}{y}_3, \alpha, \beta, \gamma \geqq 0\}$$

for the partition $P_1 = \{\{1\}, \{2\}, \{3\}\}$ which implies the constraint

$$a(g_1 + g_2 + g_3) + x_1 + x_2 + x_3 = w_1 + w_2 + w_3;$$

$$Y(P_2) = \{y_i \mid y_i = \alpha \overset{2}{y}_1 + \beta \overset{1}{y}_3, \alpha, \beta \geqq 0\}$$

for the partition $P_2 = \{\{12\} \ \{3\}\}$ which implies the constraint

$$b(g_{12}) + ag_3 + x_1 + x_2 + x_3 = w_1 + w_2 + w_3;$$

not assumed to be given; the aggregate technology set implicit in that analysis admitted the use of production technologies corresponding to any of the possible partitions. Suppose we assume that the aggregate technology set is the convex cone generated by the halflines associated with the coalition production vectors (14), (15), and (16); i.e.,

$$(17)\quad Y = \sum_{i=1}^{3} (\overset{1}{y_i}) + \sum_{i=1}^{3} (\overset{2}{y_i}) + (\overset{3}{y_1})$$
$$= conv\left(\bigcup_{k=1}^{5} Y(P_k) \right)$$

where (y_i^j) is the halfline associated with y_i^j and *conv* denotes the convex hull. Since Y is a convex cone and we have adopted Foley's other assumptions on the aggregate technology set, individual consumption sets and individual preference orderings, the Lindahl equilibrium (as defined above) will exist, it will be Pareto optimal, it will be in the core and, hence, the core will necessarily be nonempty.[21] But in Section II we produced examples where the Lindahl equilibrium was not in the core and where the core was empty. What is the reason for this contradiction?

To find the source of contradiction, consider our three-person economy with $a=b=1$, $c=3/2$, $w_1=w_2=2$, and $w_3=1$. We will first calculate explicitly the Lindahl equilibrium for this case. Observe that every allocation in Y lies within the convex cone generated by the halflines associated with the production vectors for two-person sharing;[22] i.e., $(y_1^2)+(y_2^2)+(y_3^2)$. This implies that the economy faces the resource constraint

$$(18)\quad (g_1 + g_2 + g_3)/2 + x_1 + x_2 + x_3 = w_1 + w_2 + w_3$$

Given the constraint (18), part (a) of the definition of a Lindahl equilibrium implies that $p_g^1 = p_g^2 = p_g^3$.[23] If $p_x = 1$, then $p_g^i = 1/2$ for $i=1, 2, 3$; the Lindahl allocation is $(g_1, g_2, g_3; z) = (2, 2, 1; -5/2)$ and the corresponding utility vector is $(u_1, u_2, u_3) = (2, 2, 1/2)$.[24] Using the values of the characteristic function given in (12), it is easily verified that this Lindahl equilibrium is in the core.[25]

However, if we follow the approach described in Section II, we find that the only utility vector in the core is (2, 2, 1/4) corresponding to the partition $\{\{12\}, \{3\}\}$.[26] To isolate the source of contradic-

---

$$Y(P_3) = \{y_i \mid y_i = \alpha \overset{2}{y_2} + \beta \overset{1}{y_2}, \alpha, \beta \geqq 0\}$$

for the partition $P_3 = \{\{13\}, \{2\}\}$ which imples the constraint

$$b(g_{13}) + ag_1 + x_1 + x_2 + x_3 = w_1 + w_2 + w_3;$$

$$Y(P_4) = \{y_i \mid y_i = \alpha \overset{2}{y_3} + \beta \overset{1}{y_1}, \alpha, \beta \geqq 0\}$$

for the partition $P_4 = \{\{23\}, \{1\}\}$ which implies the constraint

$$b(g_{23}) + ag_1 + x_1 + x_2 + x_3 = w_1 + w_2 + w_3; \text{ and}$$

$$Y(P_5) = \{y_i \mid y_i = \alpha \overset{3}{y_1}, \alpha \geqq 0\}$$

for the partition $P_5 = \{\{123\}\}$ which implies the constraint

$$cg + x_1 + x_2 + x_3 = w_1 + w_2 + w_3.$$

[21] The proofs in Foley (1970) are easily extended; to conserve on the length of this article, they are left to the reader.

[22] Provided that at least two of the three individuals consume positive amounts of the public good. The assumptions in our example rule out corner solutions where some individuals consume none of the public good. Observe that when $g_1 = g_2 = g_3$, the constraint (18) is the same as the constraint faced when the technology vector for three-person sharing, y_1^3, is used.

[23] Assuming that every individual consumes a positive amount of the public good.

[24] p_g^i = marginal cost = $\frac{1}{2}$. From the demand side, $p_g^i g_i = p_x x_i = x_i = w_i/2$ if $p_x = 1$ (see fn. 2) so $(g_1, g_2, g_3) = (2, 2, 1)$ and $z = -(w_1+w_2+w_3)/2 = -5/2$.

[25] The value for $v(\overline{123})$ is no longer correct since c now equals 3/2.

[26] If consumption of the public good is shared by all three individuals, then $v(\overline{123}) = 25/6$. Using this value in place of the value used for $v(\overline{123})$ in (12), we find that the game described is no longer superadditive. As discussed in fn. 7, this nonsuperadditivity implies that equation (9) no longer represents the entire set of individually rational and Pareto optimal allocations for the coalition of the whole. In particular, the sole core allocation is obtained via the partition $\{\{12\}, \{3\}\}$ for which the Pareto optimal utility frontier is given by

$$(u_1 + u_2)^{1/2} + (u_3)^{1/2} = 5/2$$

tion, examine the pattern of sharing used to attain the Lindahl allocation:

$$(19)\quad (2, 2, 1; -5/2) = (3/2)y_1^2 + (1/2)y_2^2 + (1/2)y_3^2$$

This is not a partition, but rather a set of *overlapping jurisdictions*: individuals 1 and 2 consume 3/2 units of the public good through mutual sharing which each supplements by sharing an additional 1/2 unit with individual 3. But in Section II we ruled out any pattern of sharing that assigned an individual to more than one jurisdiction; only partitions were regarded as admissible. Although the aggregate technology set admits any production vector that belongs to the union of the partition technology sets, linear combinations of these vectors will be feasible only if the jurisdictions form a partition. As a consequence, the aggregate technology set is nonconvex:

$$Y = \bigcup_{P_k \in \mathcal{P}} Y(P_k) \neq conv \bigcup_{P_k \in \mathcal{P}} Y(P_k)$$

IV. Indivisibility

The basis for this nonconvexity is, in a fundamental sense, an indivisibility—an indivisibility not of plants or commodities but of individuals. In assigning consumers to "neighborhoods," for example, we rule out the possibility of fractional assignments; a person cannot be two places at the same time. Referring to our numerical example, individuals 1 and 2 will not, for they cannot, receive two units of the public good by combining 3/2 units from their own two-person neighborhood with 1/2 unit received from two-person neighborhoods formed with individual 3. If 1 and 2 admit individual 3 as a neighbor, then a three-person neighborhood is formed, not a set of three overlapping two-person neighborhoods. But in moving from a two to a three-person neighborhood, the resources required to maintain the same "neighborhood quality" (g) increase by 50 percent. If individual 3 were as wealthy as 1 or 2, then a three-person neighborhood would be formed and the core allocation would give two units of the public good to everyone. However, individual 3 is not as wealthy as 1 or 2, and the best that a three-person neighborhood has to offer is 5/3 units of the public good.[27] Thus, two separate neighborhoods are formed: a high income neighborhood for 1 and 2 and a low income neighborhood for 3. And that outcome, drawing upon casual observation, seems a reasonable picture of what would in fact happen.

The impossibility of assigning fractions of individuals to jurisdictions is characteristic not only of neighborhoods but local public goods in general. We do not observe consumers belonging to more than one school district: the school services from several districts cannot be added to yield greater benefit. Interjurisdictional spillovers may present an opportunity for fractional membership in some sense, but there is certainly no reason to think that such spillovers will in general lead to convexity of the aggregate technology set. Although further elaboration of the model to incorporate spillovers may be worthwhile, the formulation as it now stands is rich enough to cover a number of interesting cases. Prohibiting overlapping memberships in jurisdictions has one distinct advantage—the barriers to convexity of the aggregate technology set in the presence of crowded public goods are thrown into bold relief.

Finally, one can view a household, jointly consuming a house or car, as a jurisdiction in our sense. We have, by assumption, ruled out membership in more than one household for any given in-

[27] Use equation (3).

dividual. In assessing the reasonableness of this assumption, it is important to remember that we have confined our discussion to economies having a *single* public good. We do not rule out membership in other jurisdictions—a poker group for the husband, a bridge club for the wife— if *other* public goods are involved.

V. Excludability

Implicit in our treatment of public goods is an assumption concerning the ability to exclude which should be made more explicit. Richard Musgrave, in particular, puts great stress on nonexcludability in consumption as a characteristic of public goods that is quite distinct from "nonrivalness" in consumption (or what I have called the absence of crowding).

Musgrave concludes that the presence or absence of crowding and the presence or absence of the ability to exclude gives rise to four different types of public good (see pp. 126–29). However, of the four possible cases, three have precisely the same core: the two cases of crowded and uncrowded public goods where exclusion is impossible and the uncrowded case with exclusion. If exclusion is impossible, then a blocking coalition S must assume that the members of the complementary coalition $\overline{S}$ will share the benefits of any public good produced by S; the set of utility vectors that S can obtain using only its own resources, $v(S)$, is the same whether the public good is crowded or not. The set of utility vectors that could be obtained if the members of $\overline{S}$ could be excluded from consumption is irrelevant when exclusion is impossible. If exclusion is possible but the public good is uncrowded, then $v(S)$ is the same as in the two cases just discussed: if exclusion of the members of $\overline{S}$ does not reduce the resource requirements for producing the public good for S, then whether $\overline{S}$ is excluded or not has no impact on the utility vectors achievable by S using only its own resources.[28]

The remaining case to be considered, where exclusion is possible and the public good is crowded, is the one most relevant to the analysis presented in this paper.[29] The core will be smaller than in the three preceding cases. Because exclusion reduces crowding, the blocking possibilities of any coalition S are enhanced. In our earlier discussion, exclusion has been assumed implicitly to be both *complete*[30] and *costless*, but incorporation of partial or costly exclusion into the analysis does not appear to raise any particular difficulties.

Thus, from the point of view of the analysis of allocations in the core, the four cases we have been discussing reduce to just two: the *pure public good* case (broadly defined to include the absence of the ability to exclude, the absence of crowding or both) and the *crowded public good* case with *exclusion*. Foley's model can be applied directly to the first case; the approach developed in this paper is directed to the second. In our model jurisdictions are permitted to exclude, but there is no exclusion within jurisdictions.[31] Within this analytical framework, the distinction between exclusion and crowding is inessential. Both concepts are incorporated into the definition of the aggregate technology set.

[28] However, in this case the ability to exclude may still be relevant to the bargaining process. The characteristic functions, as we have defined them, assume that the complementary coalition will "do its worst": $\overline{S}$ is assumed to provide no public good, or, if it does, to exclude S. This approach essentially rules out the "free rider" problem. Introducing "more realistic" assumptions about the complementary coalition's behavior would increase the number of allocations that can be blocked and decrease the size of core. For a related discussion, see R. W. Rosenthal.

[29] It is instructive to work through all four cases in terms of the example presented in Sections II and III.

[30] In other words, S can exclude $\overline{S}$ from all benefits of the public good produced by S.

[31] The absence of internal exclusion is essentially the defining characteristic of a jurisdiction.

VI. Prices and the Organization of Collective Consumption

I have argued that a reasonable theory of crowded public goods must recognize the nonconvexity of the aggregate technology set. Only in this way can the theory permit the existence of distinct sharing groups: neighborhoods, local political jurisdictions, clubs, and households. But in relinquishing convexity, have we given up too much? If most consumption involves sharing to some extent—within the household if nowhere else—and if nonconvexity vitiates competitive equilibrium analysis, then we are forced to conclude that the main body of economic theory has almost no relevance to the way resources are allocated.

The situation we face is remarkably parallel to that addressed by Samuelson as he neared the end of his paper on social indifference curves:

> Now I have proved the impossibility of group or community preference curves. But haven't I in a sense proved too much. Who after all is the consumer in the theory of consumer's (not consumers') behavior. Is he a bachelor? A spinster? . . In most of the cultures actually studied by modern economists the fundamental unit on the demand side is clearly the "family," and this consists of a single individual in but a fraction of the total cases.
>
> [1956, pp. 8–9]

Samuelson recognizes explicitly that family decision making raises issues "exactly of the same logical character" (1956, p. 10) as the theory of public expenditure, but his reconciliation between family decision making and standard demand analysis is limited to the special case where each family member consumes privately and the family as a whole maximizes its social welfare function. However, it is crucial for our purposes to face squarely the public character of family consumption. Fortunately, a reconciliation between "public" consumption within families and standard competitive analysis is possible—without the contrivance of a family welfare function.

Expecting the tools of competitive equilibrium analysis to explain the pattern of sharing that will emerge in an economy is clearly asking too much. General equilibrium analysis has never tried to explain household formation—or the formation of firms, for that matter. The organization of the economy, the set of consuming and producing agents, has always been regarded as given. Competitive analysis explains the economic activity (consumption and production) of these agents, not their existence.

Therefore, we might expect that competitive analysis, or some variant of competitive analysis, could be applied to an economy with public goods when the partition, the pattern of sharing, is predetermined. And that, in fact, is precisely the case. If the partition P_k is given, the assumption that $Y(P_k)$ is a convex cone is no less (and no more) reasonable than the corresponding assumption for a pure private good economy. Along with the other assumptions concerning preferences and production used by Foley, the assumption that $Y(P_k)$ is a convex cone is sufficient to guarantee the existence, Pareto optimality and inclusion in the core of the Lindahl equilibrium as we have defined it.[32] Competitive equilibrium (for pure private goods) and Foley's definition of the Lindahl equilibrium (for pure public goods) correspond to the special cases where the partition is P_1 or P_s. It is important to note that the Lindahl equilibirum is defined with respect to a given partition. In every case, Pareto optimality and inclusion in the core is proved only *relative to a partition*: there is no guarantee

[32] Modification of Foley's proofs, an elementary exercise, is left to the reader.

that the Lindahl equilibrium for some P_k will not be Pareto inferior or blocked by some allocation feasible under a different partition.

In the example used to establish Proposition 1, the Lindahl allocation corresponding to the partition $\{123\}$ was not in the core even though the only allocations in the core involve sharing of the public good by all three members of the economy.[33] It is natural to ask whether there is any other concept of equilibrium which is capable of sustaining allocations in the core. Fortunately, Foley has provided another concept of equilibrium which is much more flexible and appealing than Lindahl equilibrium, the public competitive equilibrium.[34] Translating his definition into our notation, we have for the pure public good case:

DEFINITION: *A pure public competitive equilibrium* is a feasible allocation $(g; x_1, \ldots, x_n)$, a price system $p=(p_g; p_x)$, and a vector of taxes $(t_1, \ldots, t_n)$ with $p_g g = \sum_{i=1}^{n} t_i$ such that:

(a)[35] $p\cdot(g; z) \geqq p\cdot(\bar{g}; \bar{z})$ for all $(\bar{g}; \bar{z}) \in Y(P_s)$;

(b) $p_x\cdot x_i = p_x\cdot w_i - t_i$ and if $(g; \bar{x}_i) \succ_i (g; x_i)$, then $p_x\cdot \bar{x}_i > p_x\cdot x_i$;

(c) there is no vector of public goods and taxes $(\bar{g}; \bar{t}_1, \ldots, \bar{t}_n)$

such that for every i there exists $\bar{x}_i$ with $(\bar{g}; \bar{x}_i) \succ_i (g; x_i)$ and $p_x\cdot\bar{x}_i \leqq p_x\cdot w_i - \bar{t}_i$.

In this definition, producers maximize profits, consumers maximize utility subject to their after-tax budget constraint, and there is no alternative level of public sector activity with taxes to pay for it that would make every consumer better off.

In extending this definition to an arbitrary partition P_k, we will also find it convenient to modify part (c) of Foley's version.[36]

DEFINITION: *A public competitive equilibrium under the partition P_k, $P_k=(J_1, \ldots, J_r)$*, is a feasible allocation $(g_1, \ldots, g_r; x_1, \ldots, x_n)$, a price system $p=(p_g^1, \ldots, p_g^r; p_x)$, and a vector of taxes $(t_1, \ldots, t_n)$ with $p_g^j g_j = \sum_{i\in J_j} t_i$ for all $j=1, \ldots, r$ such that:

(a) $\sum_{j=1}^{r} p_g^j g_j + p_x\cdot z \geqq \sum_{j=1}^{r} p_g^j \bar{g}_j + p_x\cdot\bar{z}$ for all $(\bar{g}_1, \ldots, \bar{g}_r; \bar{z}) \in Y(P_k)$;

(b) $p_x\cdot x_i = p_x\cdot w_i - t_i$ and if $(g_j; \bar{x}_i) \succ_i (g_j; x_i)$, then $p_x\cdot\bar{x}_i > p_x\cdot x_i$;

(c) for any $J_j, j=1, \ldots, r$, if $(\bar{g}_j; \bar{x}_i) \succ_i (g_j; x_i)$ for all $i \in J_j$,

then

$$p_g^j \bar{g}_j + \sum_{i\in J_j} p_x\cdot\bar{x}_i > p_g^j g_j + \sum_{i\in J_j} p_x\cdot x_i = \sum_{i\in J_j} p_x\cdot w_i$$

Conditions (a) and (b) have the same interpretation as the corresponding parts of Foley's definition; the budget of each jurisdiction is balanced; and if the members of a jurisdiction unanimously prefer some other allocation to the one they receive, then they cannot afford it at the given prices.[37] Foley's proofs are easily

[33] The Lindahl allocation is in the core only relative to the partition $\{123\}$: i.e., when blocking coalitions are unable to exclude nonmembers from consumption of the public good. It is worth emphasizing that the Lindahl equilibrium referred to in this example is precisely the Lindahl equilibrium as defined by Foley.

[34] Although mentioned in Foley (1970), a much more complete discussion of public competitive equilibria is contained in Foley (1967).

[35] Recall that we defined

$$z \equiv \sum_{i=1}^{n} (x_i - w_i)$$

[36] In the pure public goods case, the two definitions are equivalent.

[37] Do not confuse these prices (where there is one price for each jurisdiction) with the Lindahl prices (where there is one price for each individual). The Lindahl equilibrium is a special case of the public competitive equilibrium.

modified to demonstrate that the public competitive equilibrium for a partition P_k is Pareto optimal[38] and that for any Pareto optimum under the partition P_k there exists a supporting price system taking the form of a public competitive equilibrium.[39] It should be emphasized again that Pareto optimality is established only relative to a partition.

Since the Lindahl equilibrium is a public competitive equilibrium, we know that at least one public competitive equilibrium exists for any partition P_k. Proofs of existence for other types of public competitive equilibria could presumably be obtained for alternative theories of how the political decision-making process operates.[40]

Any imputation in the core can be sustained by some public competitive equilibrium. In the example used to establish Proposition 1, the core imputation $(u_1, u_2, u_3) = (2, 2, 11/16)$ can be achieved by a public competitive equilibrium with taxes $(t_1, t_2, t_3) = (14/15, 14/15, 19/30)$ and prices $p_x = 1$ and $p_g = 4/3$. The core imputation $(u_1, u_2, u_3) = (2, 39/16, 1/4)$ can be achieved by a public competitive equilibrium with taxes $(t_1, t_2, t_3) = (14/15, 21/30, 13/15)$ and prices $p_x = 1$ and $p_g = 4/3$.[41] Thus, the concept of a public competitive equilibrium offers much greater flexibility than the Lindahl equilibrium in adapting to the blocking power of coalitions.

The case of households treated as jurisdictions provides a useful illustration of the concept of a public competitive equilibrium. All that we have specified about the family decision-making process is that each family member maximizes his utility from goods consumed privately subject to the constraint of his "allowance" and that there is no combination of private and shared goods unanimously preferred to the bundle presently consumed by the family that it can afford. We have demonstrated that if the membership in households is considered exogenous to the model,[42] then public competitive equilibrium analysis can be applied. We have not reached a full reconciliation with standard competitive analysis, however. Consider the example of automobiles shared by members of a family. The definition of public competitive equilibrium permits a separate automobile price, p_g^i, for each household purchasing a car, but most economists would conclude that these prices will all be equal. It is reasonable to suppose, in this instance, that the resource cost of supplying a car to a family is independent of family size. In that case, condition (a) of the definition of a public competitive equilibrium implies that, for all families purchasing a car, the price must be equal; if prices were unequal, then selling all of the cars to the household having the highest p_g^i would be both feasible and more profitable. In this public competitive equilibrium, the marginal rates of substitution summed over the members of each family is equated to the marginal rate of transformation for the "public" good. Thus, the concept of a public competitive equilibrium provides a resolution

[38] See the proof in Foley (1967), pp. 60–61.

[39] See the proof in Foley (1970), pp. 68–69. As usual lump sum transfers may be required.

[40] Foley, for example, proves the existence of a public competitive equilibrium in the pure public goods case when taxes take the form of a proportional income tax (Foley (1967), pp. 71–72). However, the tax in the example used to establish Proposition 1 is a proportional income tax in the sense of Foley, so we also have a demonstration that proportional income taxes need not be in the core when public goods are crowded.

[41] The reader can verify that any imputation in the core for this example can be sustained by a public competitive equilibrium; in every case the price of the public good, p_g, equals the marginal cost of producing the public good.

[42] For example, we may assume that household membership is determined by social and cultural factors outside the scope of economic analysis.

to the problem raised by Samuelson in his paper on social indifference curves without resort to a family welfare function.

Samuelson has argued that a Lindahl equilibrium has "*no* relevance to motivated market behavior" (1969, p. 106). The fact that a Lindahl equilibrium may lie outside the core suggests that its political relevance is doubtful as well. The existence of a set of public competitive equilibria which support allocations in the core means that one does not have to rely on Lindahl equilibria to find a linkage between the allocation of public goods and a system of prices.

However, it is important to emphasize that a public competitive equilibrium has little value in itself as a solution concept.[43] In the case of a pure public good, for example, the set of public competitive equilibria coincides with the entire set of Pareto optima under that partition. To obtain a more determinate solution will, in general, require additional theory concerning the political process. There is little reason to believe that such processes bear any relation to Lindahl equilibrium.

VII. Global Lindahl Equilibrium

The two concepts of equilibrium discussed in this paper, the Lindahl equilibrium and the public competitive equilibrium, have been defined relative to the aggregate technology set, $Y(P_k)$, corresponding to a given partition P_k. As a consequence, Pareto optimality and inclusion in the core of the equilibrium allocations are proved only relative to the set of allocations achievable under P_k. There is no guarantee that such allocations will be Pareto optimal or in the core when alternative partitions are regarded as admissible.

However, if we were able to establish the existence of a *global* Lindahl equilibrium (i.e., a Lindahl equilibrium defined relative to the aggregate technology set

$$Y = \bigcup_{P_k \in \mathcal{P}} Y(P_k))$$

then stronger results would be available.

THEOREM 1: *If* $(g_1, \ldots, g_n; x_1, \ldots, x_n; p_g^1, \ldots, p_g^n; p_x)$ *is a global Lindahl equilibrium with respect to* w, *then it is in the core with respect to* w *and the aggregate technology set* Y.

PROOF:

Suppose that the coalition S could block $(g_1, \ldots, g_n; x_1, \ldots, x_n)$ by $(\bar{g}_1, \ldots, \bar{g}_n; \bar{x}_1, \ldots, \bar{x}_n)$. Since $(\bar{g}_i; \bar{x}_i) \succ_i (g_i; x_i)$ for all $i \in S$, the definition of a Lindahl equilibrium implies that

$$\sum_{i\in S} \overset{i}{p_g} \cdot \bar{g}_i + p_x \cdot \sum_{i \in S} \bar{x}_i > \sum_{i\in S} \overset{i}{p_g} \cdot g_i + p_x \cdot \sum_{i \in S} x_i = p_x \cdot \sum_{i\in S} w_i \tag{20}$$

or

$$\sum_{i\in S} \overset{i}{p_g} \cdot \bar{g}_i + p_x \cdot \sum_{i\in S} (\bar{x}_i - w_i) > 0 \tag{21}$$

But profit maximization, condition (a) in the definition of a Lindahl equilibrium, requires that

$$\sum_{i=1}^{n} \overset{i}{p_g} \cdot \bar{g}_i + p_x \cdot \bar{z} \leq 0 \quad \text{for all } (\bar{g}_1, \ldots, \bar{g}_n; \bar{z}) \in Y \tag{22}$$

This contradiction establishes the theorem. Letting $S = N$, it follows that the global Lindahl equilibrium is also globally Pareto optimal: i.e., Pareto optimal relative to the set of all allocations $(g_1, \ldots, g_n; z) \in Y$.

If a global Lindahl equilibrium exists, then much of our discussion in this paper would be superfluous; prices would guide the choice of partition as well as the

[43] I am grateful to a referee for stressing this point.

choice of an allocation under a given partition. Although nonconvexity of the aggregate technology set Y raises some doubts concerning existence of a global Lindahl equilibrium, convexity is not a necessary condition for existence. Is there something in the structure of our problem which will ensure existence of a global Lindahl equilibrium in the general case? Returning to our three-person economy, we can readily establish that this conjecture is false.

In the first place, if the core of the economy is empty, then no global Lindahl equilibrium can exist, and, in the course of demonstrating Proposition 2, we produced such a case. Furthermore, a global Lindahl equilibrium may fail to exist even when the core is not empty. Consider the example discussed in Section III where $w_1=w_2=2$, $w_3=1$, $a=b=1$, and $c=3/2$. The only allocation in the core is $(g_1, g_2, g_3; x_1, x_2, x_3)=(2, 2, 1/2; 1, 1, 1/2)$ corresponding to the partition $\{\{12\}, \{3\}\}$. This allocation is the Lindahl equilibrium relative to that partition with prices $p=(p_g^1, p_g^2, p_g^3; p_x)=(1/2, 1/2, 1; 1)$ where we arbitrarily set $p_x=1$. Note that

$$(23)\quad p\cdot(g_1, g_2, g_3; z)=p\cdot(2, 2, 1/2; -5/2)=0$$

For this equilibrium to be a *global* Lindahl equilibrium, we require

$$(24)\quad p\cdot(\bar{g}_1, \bar{g}_2, \bar{g}_3; \bar{z}) \leq 0 \quad \text{for all } (\bar{g}_1, \bar{g}_2, \bar{g}_3; \bar{z}) \in Y = \bigcup_{P_k\in\mathcal{P}} Y(P_k)$$

But $(\bar{g}_1, \bar{g}_2, \bar{g}_3; \bar{z})=(1/2, 2, 2; -5/2)\in Y$, where the allocation is achieved via the partition $\{\{1\}, \{23\}\}$, and

$$(25)\quad p\cdot(1/2, 2, 2; -5/2) = 3/4 > 0$$

Thus, no global Lindahl equilibrium exists which will sustain the only allocation in the core. We must conclude that, in general, no price system will exist to guide the choice of membership in jurisdictions, and, for that reason, we have defined our two concepts of equilibrium only relative to a given assignment of individuals into jurisdictions.

There is an interesting parallel between these results and the analysis of indivisibilities presented some time ago by Tjalling Koopmans and Martin Beckmann. They found that if the profitability of alternative locations for a given plant is independent of the location decisions of the other plants, then a system of location rents exists which supports a nonfractional assignment of plants to locations. But when profitability of alternative locations depends on the location decisions of the other plants—in their model, because of transportation costs—then no price system exists to guide the process of plant location. Prices have a limited role in explaining the organization of economic activity.

Koopmans and Beckmann stressed the importance of indivisibility to location theory in general:

> Finally, the theory of the location of economic activities has no chance of explaining such interesting realities as large and small cities without recognizing indivisibilities in the processes of production and human existence. [p. 53]

Neglect of indivisibilities, or nonconvexities, explains why the application of general equilibrium analysis to location theory has not been very fruitful. By presupposing the existence of a price system and well-defined demand or supply functions at each location, the location problem is essentially assumed away. Locations are implicitly fixed. It is for this reason that location theorists find it necessary to turn to vague notions of agglomeration or urbanization economies when they try to explain the phenomena they observe. Location theory cannot be brought within the province of general equilibrium analysis solely by adding a subscript to the conventional competitive model.

VIII. Buchanan's Theory of Clubs

A reader familiar with the work of Buchanan will by now recognize that the treatment of public goods in this paper is closely related to his theory of clubs. What I have defined as a jurisdiction, Buchanan calls a club. It is appropriate, at this point, to indicate the relationship between our respective approaches.

In the discussion so far, I have permitted crowding in the *production* of public goods but excluded the effects of crowding from individual utility functions. Specifically, I have assumed that individuals are indifferent concerning the number of other consumers sharing consumption of the public good. Buchanan, on the other hand, allows crowding to affect utility functions as well as production functions. My reason for avoiding this complication has been solely for purposes of simplifying the exposition. The model is easily extended to incorporate this additional element of crowding.

Introducing a crowding parameter, the number of individuals in a jurisdiction, into utility functions requires no change in the definition of a Lindahl or a public competitive equilibrium for a given partition P_k. Since jurisdictional size is by definition fixed, the effect of this parameter is incorporated into the preference orderings. In analyzing choice among partitions, the presence of crowding in utility functions means that the appropriate preference orderings to consider depend on the partition. Our conclusion that prices have little relevance to the selection of a partition is, if anything, reinforced by the dependence of individual preference orderings on the partition chosen.

Permitting crowding to affect preference orderings does enhance the realism of the model. Turning to an example used before, it is reasonable to expect that, holding wealth constant, larger households will purchase more cars. If a household is wealthy enough and sensitive enough to crowding, it may no longer be the preferred jurisdiction for automobile consumption: husband and wife will have their "own" car and the teenage son will use summer earnings to buy a hot rod.

From our perspective, the theory of clubs emerges as even more general than Buchanan is willing to claim. He asserts that:

> The procedure implies that the individual remains indifferent as to which of his neighbors or fellow citizens join him in such arrangements. In other words, no attempt has been made to allow for personal selectivity or discrimination in the models. To incorporate this element, which is no doubt important in many instances, would introduce a wholly new dimension into the analysis, and additional tools to those employed here would be required. [p. 13, fn. 1]

Our analysis demonstrates that introducing preference for particular individuals as members of one's jurisdiction—on the basis of race, religion or sex—is not essentially different from the introduction of crowding into the analysis.[44]

There is one area of substantial disagreement between the approaches of Buchanan and myself. He states as a necessary condition for Pareto optimality ". . . that the marginal rate of substitution 'in consumption' between the size of the group sharing in the use of X_j and the numeraire good, X_r, must be equal to the marginal rate of substitution 'in production' " (p. 5). Our analysis indicates, however, that determining the "optimal" partition—whether defined as Pareto optimal or in the core—requires *global* rather than marginal comparison of alternative sharing patterns. To any arbitrarily chosen partition there corresponds a price system which supports a public competitive equilibrium. The equi-

[44] Of course, in particular extensions of this analysis the assumption that preferences depend only upon crowding and the wealth and tastes of the other members may be a useful simplification.

librium allocation is Pareto optimal with respect to the set of allocations feasible under the given partition. However, under some other partition allocations may exist which are Pareto superior to the public competitive allocation or which permit some coalition to block. The presence of nonconvexity means that the price system is no longer a reliable guide to the attainment of allocations that are Pareto optimal and in the core. Decision making cannot be completely decentralized to the level of individual agents responding marginally to market prices. The search for the optimal pattern of sharing of public goods requires the global comparison of the allocations achievable under each alternative partition.

REFERENCES

J. M. Buchanan, "An Economic Theory of Clubs," *Economica*, Feb. 1965, *32*, 1–14.

D. Foley, "Resource Allocation and the Public Sector," *Yale Econ. Essays*, spring 1967, *7*, 43–98.

———, "Lindahl's Solution and the Core of an Economy With Public Goods," *Econometrica*, Jan. 1970, *38*, 66–72.

L. Johansen, "Some Notes on the Lindahl Theory of Determination of Public Expenditures," *Int. Econ. Rev.*, Sept. 1963, *4*, 346–58.

T. C. Koopmans and M. Beckmann, "Assignment Problems and the Location of Economic Activities," *Econometrica*, Jan. 1957, *25*, 53–76.

E. Lindahl, "Just Taxation—A Positive Solution," reprinted in part in R. A. Musgrave and A. T. Peacock, eds., *Classics in the Theory of Public Finance*, London 1958.

R. A. Musgrave, "Provision for Social Goods," in J. Margolis and H. Guitton, eds., *Public Economics*, New York 1969, 124–44.

J. M. Ostroy, "A Suggestion for Complicating the Theory of Exchange," paper presented to the meetings of the Econometric Society, Dec. 1969.

R. W. Rosenthal, "External Economies and Cores," *J. Econ. Theor.*, June 1971, *3*, 182–87.

P. A. Samuelson, "The Pure Theory of Public Expenditures," *Rev. Econ. Statist.*, Nov. 1954, *36*, 387–89.

———, "Social Indifference Curves," *Quart. J. Econ.*, Feb. 1956, *70*, 1–22.

———, "Pure Theory of Public Expenditure and Taxation," in J. Margolis and H. Guitton, eds., *Public Economics*, New York 1969, 98–123.

H. Scarf, "The Core of a N-Person Game," *Econometrica*, Jan. 1967, *35*, 50–69.

——— **and G. Debreu**, "A Limit Theorem on the Core of the Economy," *Int. Econ. Rev.*, Sept. 1963, *4*, 235–46.

L. S. Shapley, "On Balanced Sets and Cores," *Naval Res. Logist. Quart.*, Dec. 1967, *14*, 453–50.

——— **and M. Shubik**, "On the Core of an Economic System with Externalities," *Amer. Econ. Rev.*, Sept. 1969, *59*, 678–84.

M. Shubik, "Edgeworth Market Games," in A. W. Tucker and R. D. Luce, eds., *Contributions to the Theory of Games*, Vol. IV, Princeton 1959, 267–78.

The Supply of Rental Housing: Comment

By RONALD E. GRIESON*

In a recent issue of this *Review*, Frank de Leeuw and Nkanta Ekanem (L-E) proposed to determine whether the marginal cost of housing is constant or rising as a function "of the amount of housing services provided in an area."

I propose to show that a nonrising supply curve is inconsistent with positive long-run land prices and to show that the L-E housing supply elasticities, calculated from reduced form equations, are very sensitive to the specification of the true supply and demand functions. An alternative and perhaps more reasonable a priori specification is herein set out which is then used to calculate much different elasticities of supply.

These alternatively calculated coefficients are elastic, rather than inelastic as deduced by L-E, and consistent with estimates which I have made in another paper, and with the reduced form estimated by the authors. The L-E model and estimates are shown to suffer from misspecification, statistical inconsistency, and multicollinearity.

I. Consistency of the Hypothesis

If more housing could be produced in a given land area at constant cost, then the existing housing stock could be constructed at equal cost on a smaller area of land. Hence land having no marginal product would earn no rent.[1] Of course the authors may not have been precise in their use of the word "area."

II. The Econometric Model

L-E postulate the following demand and supply equations for housing, (12) and (13),[2] respectively:

$$(S - H) = \alpha_1 + \beta_1 Y - \beta_2(R - P) \tag{12}$$

$$R = \alpha_2 + \beta_3 C + \beta_4 O + \beta_5(S - H) + \beta_6 H \tag{13}$$

All variables measured in logarithmic form:

S = the quantity of housing services
H = the number of households
Y = real income
R = rents per unit of housing services
P = the price level of nonhousing goods and services
C = the price of capital inputs
O = the price of operating inputs

The authors cite a study by de Leeuw in which the income and price elasticities of rental housing are calculated as $+1.0$ and -1.0, respectively. They argue that with minor qualification this permits β_1 and β_2 to be set (approximately) equal to 1.0. Setting β_1 and β_2 equal to 1.0 and solving (12) and (13) for R leads to the reduced form:

$$R = \left(\frac{\alpha_2 + \alpha_1\beta_5}{1+\beta_5}\right) + \left(\frac{\beta_3}{1+\beta_5}\right)C + \left(\frac{\beta_4}{1+\beta_5}\right)O + \left(\frac{\beta_5}{1+\beta_5}\right)(Y+P) + \left(\frac{\beta_6}{1+\beta_5}\right)H \tag{14}$$

They then proceed to estimate (14) for low income housing, R_1, moderate income housing, R_2, and high income housing, R_3. For example,

$$\underset{}{R_1} = \underset{(2.0)}{2.33} + \underset{(3.9)}{.321C} + \underset{(2.1)}{.183\,O} + \underset{(4.0)}{.593(Y+P)} - \underset{(-2.6)}{.036H} \qquad (R^2 = .70) \tag{15}$$

They calculate $\beta_5 = 1.46$ and the elasticity of supply of housing equal to $(1/\beta_5) = 0.7$

* Assistant professor of economics, Queens College, City University of New York.

[1] For a fuller explanation of this point, see Grieson.

[2] Equation numbers are those used by de Leeuw and Ekanem except for primed numbers which are mine.

TABLE 1—ALTERNATIVELY CALCULATED PARAMETERS

Equation	β_3	β_4	β_5	Elasticity of supply $=1/\beta_5$	A priori coefficient of $P=(1-\beta_3-\beta_4+\beta_5)/(1+\beta_5)$	Actual coefficient of P, t-ratio below
(9) low income housing, R_1	.21	.14	.45	2.2	.76	.89 (2.1)
(10) moderate income, R_2	.19	.08	.54	1.8	.83	1.00 (2.0)
(11) high income, R_3	.53	−.37	2.7	.37	.96	1.26 (1.7)

(0.5 and 0.3 for R_2 and R_3, respectively). These values are very low and much below those Muth and I have reported.

Using the same notation, let us now postulate the following alternate set of demand and supply equations, which can be estimated from reduced forms calculated by L-E and which demonstrates the sensitivity of their results to changes in a priori specification.

$$(1') \quad (S - H) = \alpha_1 + \beta_1 Y - \beta_2(R - P)$$

$$(2') \quad R - P = \alpha_2 + \beta_3(C - P) + \beta_4(O - P) + \beta_5(S - H)$$

These equations seem more reasonable since there is no reason to include H in the supply equation as L-E did, obtaining difficult to explain negative coefficients. Furthermore R and perhaps C, and O ought to be deflated by P à la the authors' deflation of R by P in the demand equation (1). Setting $\beta_1=\beta_2=1$ again and solving (1′) and (2′) for R we obtain:

$$(3') \quad R=\frac{\alpha_2+\alpha_1\beta_5}{1+\beta_5}+\frac{\beta_3 C}{1+\beta_5}+\frac{\beta_4 O}{1+\beta_5}+\frac{\beta_5 Y}{1+\beta_5} + \frac{(1-\beta_3-\beta_4+\beta_5)}{(1+\beta_5)}P$$

which would be estimated as:

$$(4') \quad R = d_i + d_{i1}P + d_{i2}Y + d_{i3}C + d_{i4}O$$

L-E have estimated this reduced form obtaining the following three equations with the t-ratio shown in parentheses below coefficients:

$$(9) \quad \underset{(4.1)}{R_1 = 4.39} + \underset{(2.1)}{.89P} + \underset{(2.4)}{.31Y} + \underset{(3.3)}{.31C} + \underset{(2.2)}{.20O}, \qquad R^2 = .66$$

$$(10) \quad \underset{(3.3)}{R_2 = 4.32} + \underset{(2.0)}{1.00P} + \underset{(2.2)}{.35Y} + \underset{(2.4)}{.27C} + \underset{(1.1)}{.12O}, \qquad R^2 = .54$$

$$(11) \quad \underset{(1.8)}{R_3 = 1.59} + \underset{(1.7)}{1.26P} + \underset{(3.1)}{.73Y} + \underset{(.8)}{.14C} - \underset{(-.6)}{.10O}, \qquad R^2 = .36$$

We can now set $d_{i2}=(\beta_5/1+\beta_5)$ finding β_5 and the elasticity of supply of housing equal to $(1/\beta_5)$. We can also solve for $\beta_3=(1+\beta_5)d_{i3}$ and $\beta_4=(1+\beta_5)d_{i4}$, from which we can obtain a priori estimates of d_{i1}, the coefficients of P, equal to $(1-\beta_3-\beta_4+\beta_5)/(1+\beta_5)$, providing an econometric test for our a priori model. Using this a priori specification and equations (9)–(11), we calculate the following alternate parameters (shown in Table 1).

A simpler alternative specification of the supply function is:

$$(5') \quad (R - P) = \alpha_2 + \beta_3 C + \beta_4 O + \beta_5(S - H)$$

Again, assuming $\beta_1=\beta_2=1$ and substituting from the demand equation (1′) yields:

$$(6') \quad R = \frac{\alpha_2 + \alpha_1\beta_5}{1+\beta_5} + \frac{\beta_3 C}{1+\beta_5} + \frac{\beta_4 O}{1+\beta_5} + \frac{\beta_5 Y}{1+\beta_5} + P$$

All coefficients except that of P stay the same while our a priori prediction of the coefficient of P becomes one. The calculated coefficients of P and the t-ratios in L-E's equations (9), (10), and (11) are, respectively, .89 ($t=2.1$), 1.00 ($t=2.0$) and 1.25 ($t=1.7$). Hence, our second alternative model also fits L-E's estimates extremely well.

The a priori assertion that the coefficient of P is equal to $(1-\beta_3-\beta_4+\beta_5)/(1+\beta_5)$ or one is held up by equations (9)–(11), but the L-E assertion that the coefficients of Y and P are the same does not seem to be. They have put H into equations (15)–(16) and defined a new variable $Y+P$, but the absence of H should not really be of great importance in determining the coefficients[3] of (9)–(11) where we can examine the hypothesis that the coefficients of P and Y are the same. L-E have not made such examination.

TABLE 2—COMPARISONS OF Y AND P COEFFICIENTS

Equation	Coefficient Y	Coefficient P
(9)	0.31 (2.4)	0.89 (2.1)
(10)	0.35 (2.2)	1.00 (2.0)
(11)	0.73 (3.1)	1.26 (1.7)

Though we cannot test whether the coefficients of P and Y are significantly different without knowing their covariance, we can say that they appear so at the 10 percent level assuming no covariance. Positive covariance as is likely would probably show them significantly different at the 5 percent level or higher.

[3] The assumption is supported by the observation that the coefficients of O and C are not significantly affected by the inclusion or exclusion of H. For example, compare (9)–(11) and (15)–(17).

The a priori model of supply which would best fit land value theory would divide the supply of housing services variable, S, by the quantity of land upon which S is located, T_s, as follows.

$$(7') \quad R - P = \alpha_1 + \beta_3 C + \beta_4 O + \beta_5 (S - T_s)$$

III. Conclusions

First we have briefly explained why recent theoretical findings show the L-E hypothesis of constant cost of increasing the amount of housing services on a given land area is inconsistent with the existence of positive land values. We have gone on to suggest the use of housing structures per unit land as a variable which should have been used to reflect the above fact.

Secondly, we pose two reasonable alternative specifications of the supply function which would yield reduced form equations of a type estimated by L-E. Using these alternative specifications we obtain estimates of the elasticities of supply of housing and a priori predict the coefficients of P found in the actual reduced form. The ability to predict the coefficients of P indicates that our alternative specification is likely to be superior to those postulated by L-E which predict the coefficients of P and Y equal in apparent contradiction of the estimated parameters.

The second alternative specification advanced in this paper predicts the coefficient of P as always being one while the first alternative specification predicts P coefficients of .76 to .96. The coefficients calculated by L-E are .89, 1.00, and 1.26; none of which is significantly different from either alternative specification and all of which are significantly different from the P coefficient calculated by L-E when they force the coefficient of P and Y to be equal.

The elasticities of supply calculated using our alternative specification and the reduced form estimators for low and moderate rental housing are 2.2 and 1.8 compared to .7 and .5, respectively, calculated by L-E. Supply elasticities of 2.36 to 5.00 are predicted in my earlier work when land is about one-fifth to one-tenth of the value of structure plus

land[4] where the elasticity of supply of structures is one-half of the ratio of total cost to land cost;

$$E_s = \frac{1}{2}\,\frac{\text{Total Value}}{\text{Land Value}}$$

The estimated equations for high income housing, (11) and (17), yield supply elasticities of .3 for L-E and .37 for the alternative specification. These estimators are the most econometrically suspect, since the coefficient of O, the *log* of the price of operating inputs, turns "negative". The t-ratios for O, C, and P drop to $-.6$, .8 and 1.7, respectively, in equation (11) from 2.2, 3.3, and 2.1 in equation (9), while the R^2 drops to .36 from .54 and .66 in equations (9), (10), and (11), respectively.

Examining the authors' specifications of the reduced form,

$$\text{(15)}\quad \underset{(2.0)}{R_1 = 2.33} + \underset{(3.9)}{.321C} + \underset{(2.1)}{.183O} + \underset{(4.0)}{.593(Y+P)} - \underset{(-2.6)}{.036H} \qquad (R^2 = .70)$$

$$\text{(17)}\quad \underset{(.6)}{R_3 = 1.43} + \underset{(1.1)}{.178C} - \underset{(-.4)}{.066O} + \underset{(2.6)}{.745(Y+P)} + \underset{(.2)}{.004H} \qquad (R^2 = .35)$$

[4] The one-fifth figure is based upon HUD data for new homes built in the United States in 1970. The one-tenth figure is based on the land content of high-rise apartment buildings.

we see the same dramatic drop in: the t-ratios of all variables; the coefficient of O to negative; and the R^2.

The explanation of this difficulty would seem to be that variables O, C, P, and perhaps Y suffer from multicollinearity in equation (11) and (17). The coefficients of C and O are dropping to below their true value as the coefficients of Y and P are rising above their true values. Thus the coefficients of P and Y are too high and the elasticity estimates, $1/\beta_5$, too low in these equations, though it would not be surprising to find the elasticity of supply of high rent housing somewhat more inelastic than that for low rent housing, since it is easier to depreciate existing housing downward than it is to upgrade low quality housing. High income housing may also have a higher component of historic, artistic, and cultural resources which are difficult to duplicate, thereby reducing supply elasticity.

This multicollinearity may have also lowered the estimates of supply elasticity for moderate and to a lesser extent low income housing. Hence, the supply elasticity of 2.2 estimated for low income housing is the most significant by all criteria and the closest to those which I estimated.

REFERENCES

F. de Leeuw and N. Ekanem, "The Supply of Rental Housing," *Amer. Econ. Rev.*, Dec. 1971, *61*, 806–17.

R. Grieson, *The Economics of Land Values and Property Taxes*, M.I.T., working pap. no. 72, June 1971.

R. Muth, *Cities and Housing*, Chicago 1969.

The Supply of Rental Housing: Reply

By Frank de Leeuw and Nkanta F. Ekanem*

Ronald Grieson comments on two aspects of our paper: the plausibility of its results, and the interpretation of its empirical findings.

On empirical findings, we have no strong preference for our interpretation over Grieson's. He shows, in essence, that if the general price level enters the supply function as well as the demand function, then our regression coefficients may imply elasticities of supply somewhat higher than the ones we presented. The elasticities we presented ranged from .3 to .7; the ones he presents range from .4 to 2.2. Both Grieson's interpretation and ours lead to the same principal conclusion; namely, that " . . . the long-run supply of housing services is less than perfectly elastic" (see de Leeuw and Ekanem, p. 812). Both sets of elasticity estimates suffer from the problems of measurement error which our article discussed.

An advantage of Grieson's interpretation is that it implies higher reduced-form coefficients for the price level than for income per household, which fits the data better than our specification. One disadvantage of his interpretation is that there is no clear theoretical basis for including the general price level in the supply function; if the supply function is the result of profit maximization subject to a production function for converting capital and operating inputs into housing services, then the only relevant prices would seem to be those of housing services, capital inputs, and operating inputs (all of which were already included in our specification). Possibly the general price level serves as a proxy for prices of inputs not covered by our measures. But surely the general price level does not belong in the supply function simply because it appears in the demand function, as Grieson seems to argue.

On the plausibility of results, we have reservations about Grieson's position. He finds a long-run rising supply price for housing services plausible because of the inelastic supply of land. Both our elasticity estimates and his, however, refer to the elasticity of supply *holding constant* the price of land as well as prices of other inputs. (See de Leeuw and Ekanem, pp. 808, 810.) Hence, the explanation of our findings cannot be the tendency for the price of land to rise as housing demand shifts upward.

We suspect that the explanation of our findings lies in diseconomies of scale in the maintenance, improvement, and conversion of existing housing capital. To understand the connection between these activities and our study it is useful to distinguish between two meanings of "long run" in the housing market. The first is a time period long enough for the total *quantity* of housing capital in a housing market to respond fully to changes in underlying conditions—a period whose length has been estimated to be of the order of magnitude of six years. (See Richard Muth, p. 76.) The second is a time period long enough for not just the quantity, but the *form* of existing housing capital—number of units per building, architectural style, location pattern within a housing market area, and so forth—to respond fully to changes in underlying conditions. The second long run is surely a great deal longer than six years.

We believe that the results of our study apply to the first rather than the second of these long runs. In the second long run, at least under competitive conditions, prices per unit of housing service presumably approach new construction prices (plus local land rents) for all types of housing. Since the construction of new housing is a replicable process, we might expect (input prices aside) constant returns to scale to characterize, at least approximately, this long run.

In the first long run, however, there could be important quasi rents—positive or nega-

* The Urban Institute.

tive—for particular types or quality levels of existing housing, arising from diseconomies of scale in altering capital which has already been embodied in a specific form and location. Since we believe that this process of altering existing capital constitutes a significant part of what we observe in our cross-section study, we feel that it may well be the source of the rising supply price we observe across metropolitan housing markets.

REFERENCES

F. de Leeuw and N. F. Ekanem, "The Supply of Rental Housing," *Amer. Econ. Rev.*, Dec. 1971, *61*, 806–17.

R. E. Grieson, "The Supply of Rental Housing: Comment," June 1973, *63*, 433–36.

R. Muth, "The Demand for Nonfarm Housing," in A Harberger, ed., *The Demand for Durable Goods*, Chicago 1960, pp. 29–96.

Property Crime and Economic Behavior: Some Empirical Results

By DAVID LAWRENCE SJOQUIST*

Belton Fleisher and Gary Becker have previously suggested that crime may be explainable, at least in part, by economic theory. This paper will consider this possibility for the crimes of robbery, burglary, and larceny over $50.[1] The hypothesis to be tested is that under some conditions, criminals can be treated as rational economic beings, assumed to behave in the same economic manner as any other individual making an economic decision under risk.

In Section I we develop an economic model of criminal behavior; Section II is devoted to a discussion of the problems involved in testing the model; and Section III is concerned with the results of the empirical analysis.

I. An Economic Model of Criminal Behavior

Given a fixed amount of time, t, an individual must choose between those various activities which consume time. For our purposes, let us assume there are two time-consuming activities or forms of behavior, one of which is legal,[2] i.e., work, and the other illegal, i.e., crimes against property. Our task is to explain how the individual allocates his time between the two activities. It is expected that this depends on the gains and costs involved in the two activities.

Participation in the two activities is expected to result in psychic and financial gains and costs. The gain per unit of time from legal activity is measured by the individual's wage rate, g_w, which is assumed to be constant and to include any psychic gain. The individual's total gain from legal activity, $\bar{g}_w$, therefore equals the time allocated to that activity, t_w, times the wage rate. We assume that g_w is generated with certainty with all associated costs both financial and psychic included.

Participation in illegal activities results in a psychic and/or financial gain. The psychic gain is measured by that quantity of money which the individual is willing to pay to obtain the psychic gain. The financial gain is measured directly by the dollar value to the criminal of the assets stolen.[3]

Since a crime could be unsuccessful, the gain from illegal activity, particularly the financial gain, is random with the individual normally possessing a subjective probability distribution of the possible gains. We will simplify however, by assuming that the gain is generated with certainty. We will also assume that the wage rate from crime, g_c, is constant and includes any psychic gain so that the total gain from illegal activity, $\bar{g}_c$, is equal to the time spent in illegal activity, t_c, times the wage rate from crime.[4] All costs of illegal activity, both financial and psychic, other than those associated with arrest, conviction, and punishment, are included in g_c.

Besides any expenditures for tools, etc., illegal activity involves the possibility of costs resulting from arrest, conviction, and punishment. These include public scandal, loss of freedom, the distaste of prison life, lawyer fees, possible reduction in potential income, since individuals possessing crim-

* Assistant professor of economics, Georgia State University. I wish to thank the National Institute of Law Enforcement and Criminal Justice for financial support of this paper, a fact which does not necessarily indicate the concurrence of the Institute in the statements or the conclusions contained herein; Edward Foster, without whose help this paper would not exist; and James Henderson, N. James Simler, Raymond Galvin, James Weatherby, Barbara Zoloth, and Larry Schroeder for many helpful comments.

[1] For a definition of these crimes, see Federal Bureau of Investigation (FBI), *Reports*, pp. 13–26.

[2] If leisure and work are equally valued at the margin then we need only consider one of them as the legal activity.

[3] Since nonfinancial assets are normally distributed through a black market, the value to the criminal will not necessarily be the same as the loss of the victim.

[4] Assume $t_w, t_c \geq 0$, $t > 0$ and $t_w + t_c = t$.

inal records find it more difficult to obtain employment, and loss of earnings while imprisoned, measured by the discounted value of income foregone while in prison less any benefits received, such as room and board and vocational training.

The psychic and financial costs associated with arrest and conviction are assumed to be quasi-fixed costs, that is, if arrested and convicted, the cost is the same regardless of the time the individual spent in illegal activity.[5]

The costs associated with imprisonment are variable and depend upon the length of the sentence served and the individual's legal wage rate. Assuming that the expected length of the sentence received is positively related to the time spent in illegal activity,[6] and that the length of the sentence received is positively related to the length of the sentence actually served, we can conclude that the expected cost of imprisonment depends upon the legal wage rate and the time spent in illegal activity. (There is a substantial amount of evidence indicating that variance in sentencing is large,[7] implying the existence of a probability distribution of possible costs from imprisonment, given that arrest and conviction have taken place. We will, however, disregard this and assume that there is no uncertainty as to the sentence received.)

We then have the total cost of illegal activity, $\bar{p}$, composed of a quasi-fixed cost component, p^*, and a variable cost component, $\hat{p}=\hat{p}\,(g_w, t_c)$. Thus, $\bar{p}=p^*+\hat{p}\,(g_w, t_c)$. Assume for simplicity that $\partial\hat{p}(g_w, t_c)/\partial t_c=p$ and is constant.

All crimes do not result in arrest and conviction, meaning that the cost of illegal activity occurs according to some probability distribution. Individuals are assumed to possess a subjective evaluation of the probabilities of arrest, of conviction, and of imprisonment. Although these probabilities no doubt depend upon the time spent in illegal activity, we will assume they are constant. Since nothing is lost in the analysis by doing so, we will also assume temporarily that the probability of conviction and punishment, conditioned on arrest, is one. Thus the joint probability of arrest, conviction, and punishment equals the probability of arrest.

The gains and costs from legal and illegal activities are evaluated by the individual according to his preference ordering, taking into account the various probabilities. We assume that this preference ordering is a Von Neumann-Morgenstern measurable utility function, that is, under conditions involving risk, individuals will choose between alternatives so as to maximize the expected value of utility.

There exist two possible states of nature, the individual will either be arrested and therefore convicted and punished, or not arrested. Let $U(\bar{g}_w+\bar{g}_c)$ represent the ordinal utility if not arrested, and $U(\bar{g}_w+\bar{g}_c-\bar{p})$ represent the ordinal utility if arrested, where the utility index is appropriately chosen in accordance with the Von Neumann-Morgenstern axioms. Let r represent the probability of arrest, conviction, and punishment; that is, the probability of incurring those financial and psychic costs associated with arrest, conviction, and punishment. The expected total utility is thus:

$$E(U) = (1-r)U(\bar{g}_w+\bar{g}_c) + rU(\bar{g}_w+\bar{g}_c-\bar{p}) \tag{1}$$

which is to be maximized subject to the constraint $t=t_w+t_c$.

The first-order conditions for maximization require that

$$\frac{g_w-g_c}{g_c-g_w-p} = \frac{(1-r)U_1'}{rU_2'} \tag{2}$$

where

$$U_1' = \frac{dU(\bar{g}_w+\bar{g}_c)}{d(\bar{g}_w+\bar{g}_c)}$$

[5] If the quasi-fixed costs are large, then an increase in the length of the sentence increases total cost by a small percent. The effect then would be for variations in the length of the sentence to have little influence on the amount of time spent in illegal activity.

[6] For a study of sentencing which yields this result, see Edward Green.

[7] See Glendon Schubert.

and

$$U_2' = \frac{dU(\bar{g}_w + \bar{g}_c - \bar{p})}{d(\bar{g}_w + \bar{g}_c - \bar{p})}$$

Assuming U_1', $U_2'>0$, then condition (2) holds if

$$g_c - g_w < p \tag{3}$$

It should be pointed out that a corner solution at $t_w = t$ is quite possible because of the quasi-fixed cost of illegal activity.

The second-order condition requires that

$$(1 - r)U_1''[g_c - g_w]^2 + rU_2''[g_c - g_w - p]^2 < 0, \tag{4}$$

where U_1'', U_2'' refer to the second derivatives. A sufficient condition for (4) to hold is that the individual is risk averse, i.e., U_1'', $U_2''<0$. If an individual has a strong enough preference for risk the indifference curves will be concave, in which case the individual will either specialize in work or crime.

The question we wish to ask is: How do people respond to changes in risk, wages, and punishment? If the individual is risk averse, then given our previous assumptions, it can be shown that $dt_w/dr>0$, i.e., an increase in r reduces the time spent in illegal activity. However if the indifference curves are concave then a person who was specializing in work will now specialize in crime, if the change in r is large enough. Likewise it can be shown that for an individual who is risk averse we would have $dt_w/dg_c<0$, $dt_w/dg_w>0$, and $dt_w/dp>0$.

II. Testing the Model

The time spent in criminal activity by the ith individual, t_c^i, can be expressed as a function of the variables specified in the previous section, namely,

$$t_c^i = f^i(r^i, p^i, g_w^i, g_c^i; x^i) \tag{5}$$

where x^i is an index of variables which "measures" tastes, and the other variables are as previously defined.

Assume that this function has the following explicit form:

$$t^i = \alpha_0 r^{i\alpha_1^i} g_w^{i\alpha_2^i} g_c^{i\alpha_3^i} p^{i\alpha_4^i} x^{i\alpha_5^i} \tag{6}$$

The model is specified in terms of individual behavior, while the available data are community aggregates. Taking the natural *logs* of (6) and replacing individual values by community means we have

$$ln\,(TC/I) = ln\,\alpha_0 + \alpha_1\,ln\,r + \alpha_2\,ln\,GW + \alpha_3\,ln\,GC + \alpha_4\,ln\,P + \alpha_5\,ln\,X \tag{7}$$

where (TC/I), GW, GC, P, and X represent population means of t_c, g_w, g_c, p, and x, respectively, TC is the total time spent in illegal activity and I is the community population. The values of the parameters of (7) will be the same as the parameters of (6) if we assume that the values of the parameters of (6) are the same for all individuals in a given community[8] and that the density function between individual values and the community mean of the variable are homogeneous of degree one with respect to changes in the individual values and community means.[9]

A cross-sectional sample of 53 municipalities with 1960 populations of 25,000 to 200,000 was selected on the basis of the following criteria:

1. We eliminated communities with 1960 populations of less than 25,000 because of a lack of data, and of more than 200,000 in order to use communities which were internally homogeneous.
2. To minimize spillover effects, any community with neighboring communities of 25,000 or more was eliminated, as well as any community with less than 50 percent of the population of its Standard Metropolitan Statistical Area (*SMSA*).

Of the 75 communities which fulfilled these criteria, 22 had to be dropped for lack of information on arrests.

Since it is not possible to directly measure the time spent in illegal activity, we measured the amount of illegal activity by the

[8] For a proof see Henri Theil, p. 142.
[9] See James Tobin, p. 126.

total number of property crimes committed, N. This in turn was measured by the total number of the three types of property crimes recorded by the local police department as reported in the 1968 FBI *Reports*.

The data on reported crimes is subject to strong criticism[10] and hence one must be very cautious in drawing implications using this data.

The FBI supplied unpublished information, collected yearly from local police departments, which we used to measure arrests and convictions at the local level for the year 1968. The number of convictions was similarly estimated[11] for 1968 from this information.[12]

We employed three separate measures of r: the ratio of arrests to the number of crimes; the ratio of convictions to the number of crimes; and the ratio of convictions to arrests. These measure, respectively, the probability of arrest, $r(A)$, the probability of conviction, $r(AC)$, and the probability of conviction given arrest, $r(C|A)$.

When the ratio of arrests or convictions to reported crimes is used to measure r, reported crimes appear in both the dependent and the independent variable. Thus to the extent that there are errors in the measure of reported crimes, we expect a bias in the coefficients of that equation.

The net gain from legal activity, GW, was measured by the annual labor income to manufacturing workers in 1968. These are county-wide data and were obtained from the 1968 *County Business Patterns*.[13]

It is not possible to differentiate the incomes of criminals from the incomes of victims, which poses a problem in measuring the gain from legal activity. To the extent that crimes against high income individuals yield greater returns to the criminal, we would expect a positive correlation to exist between income and the number of crimes. If this is true then use of mean community income is not a good measure of the opportunity cost of committing crime. In addition to mean income therefore, we also employed substitute variables: the 1968 labor force unemployment rate, derived from information published by the Bureau of Labor Statistics, and the percent of families with income below \$3,000, taken from the 1960 *Census of Population*. An unemployed person has more time to allocate to illegal activity and since his current income is low, he would have a greater incentive to commit crimes. The same reasoning can be applied to a family with a low income.

We have been unable to find any information on the financial loss of the public from criminal activity, other than a national average for each classification of crime. As a substitute variable we used retail sales per establishment reasoning that the size of the expected gain depends upon the size of the establishment a criminal chooses to rob or burglarize.

We assumed that the only variable cost of illegal activity is the cost of punishment, which depends upon the income foregone while in prison less benefits, and any psychic costs. The discounted value of the foregone earnings is highly correlated with current earnings thus we cannot use both in our regression equation. Given the mean income for the community, the foregone earnings will be proportional to the length of the sentence served. Thus the length of the sentence served can be used to reflect the cost of illegal activity. (Since we had no measure of the psychic costs, we had to ignore them.)

[10] For a discussion of the problems with crime data see Thorsten Sellin and Marvin E. Wolfgang, pp. 71–86, and The National Opinion Research Center *Report*.

[11] Convictions were determined in the following manner: A conviction rate was found using the ratio of persons guilty to the sum of persons guilty and acquitted. This ratio was then applied to those cases not yet brought to court. The number of convictions was then the sum of this product, the number guilty and the number sent to juvenile court.

[12] Using the number of arrests and convictions and the number of crimes for the same period does not take into account time lag between the criminal act and the arrest and the resulting conviction. This may result in erroneous measures of r.

[13] The size of the city and the method of selection of the sample resulted in a sample for which differences between the city and the rest of the county was minimized, so that the problem of using county-wide data is minimal.

To measure the average sentence served, we used the reported state-wide average time served by inmates released from state and federal prisons in 1960, for the crimes we are considering, as published by the Federal Bureau of Prisons. Figures other than state-wide averages or beyond 1960 were not available.[14]

A number of variables was employed to reflect economic and demographic differences between communities as well as to reflect possible differences in the evaluation of the psychic gains and costs. The percent of the population which is nonwhite reflects the income distribution and regional differences, since in our sample the percent nonwhite decreases as one moves from the South to the Northeast, to the West, to the Midwest. The mean number of school years completed reflects both cultural differences and differences in expectations of future incomes. We employed population density to reflect the fact that the closer the opportunities are, the greater the likelihood the individual can acquire the information necessary to perform a crime. Finally, we included population as a factor because of possible differences in the social make-up of residents of different sized cities.

Thus we arrive at the equation we estimated:

$$(8)\quad (N/I) = \beta_0 r^{\beta_1} S^{\beta_2} GW^{\beta_3} E^{\beta_4} NW^{\beta_5} SY^{\beta_6} D^{\beta_7} I^{\beta_8} \phi$$

where N, r, GW, and I are as previously defined, S is the average prison sentence served, E is retail sales per establishment, NW is the percent nonwhite, SY is the mean school years completed, D is population density, and ϕ is a random error term assumed to have a lognormal distribution.

[14] We feel the assumption that the same sentence is served by all criminals for a particular crime in a given state regardless of the community is not unreasonable. We know that the variance in the sentence served across states is substantially less than the variance in the sentence imposed. In addition, since paroles are normally granted by one agency for all state prisons in a particular state, we expect even more uniformity in the sentence served than in the sentence imposed.

III. Results of the Statistical Analysis

Using multiple regression analysis, we estimated the parameters of equation (8) using the total of the reported property crimes in the three categories, robbery, burglary, and larceny over \$50 to measure N and employing three different measures of risk: $r(A)$, $r(AC)$, and $r(C|A)$. (We are assuming that if convicted the individual is punished.) The results of the regressions are presented in Table 1.

It was implied from our model that the coefficients of the measure of risk should be negative and as seen in Table 1 the coefficients are of that sign. However, when r is measured by $r(A)$ and $r(AC)$ the coefficients are biased, as indicated above. Note though, that the coefficient for convictions per arrest, which is not biased, is also negative.

According to the theory, an increase in r reduces the amount of crime because it increases the expected cost and therefore lowers expected utility from crime. Separate increases in the three measures of risk result in increases in different costs. We thus expect the elasticity of N with respect to risk to be higher for $r(AC)$ than for $r(A)$ given $r(C|A)$, and for $r(C|A)$ given $r(A)$. It is also expected that the elasticity for $r(C|A)$ would be less than the elasticity for $r(A)$, given our discussion of cost in Section I.

Equation 4) of Table 1 presents the results of a regression in which both $r(C|A)$ and $r(A)$ were used. Comparing equation 4) with equation 2), we see that the elasticity for $r(C|A)$ is smaller than the elasticity for $r(AC)$ as expected, though not significantly. The elasticity for $r(A)$, however, is slightly larger than the elasticity for $r(AC)$. We note also from equation 4) that elasticity for $r(C|A)$ is less than the elasticity for $r(A)$, but again not significantly.

The coefficient for average sentence served, S, is negative for all equations in Table 1, a result our theory leads us to expect.

In equation 4), we note that the elasticity of N/I is greater with respect to $r(A)$ than with respect to S, but not significantly. If we assume risk aversion, i.e., that the ex-

TABLE 1—REGRESSION EQUATIONS USING N TO MEASURE NUMBER OF CRIMES
(t-statistics in parentheses)

1) $$\ln N/I = -1.09 - .342 \ln r(A) - .212 \ln S + .167 \ln E + .142 \ln GW + .179 \ln I$$
$$(-3.374)^{*} \quad (-1.070) \quad (.547) \quad (.526) \quad (1.402)^{*}$$

$$+ .126 \ln NW + .702 \ln SY + .051 \ln D$$
$$(3.627)^{*} \quad (1.255) \quad (.438) \qquad R^2 = .506$$

2) $$\ln N/I = -1.46 - .354 \ln r(AC) - .292 \ln S + .071 \ln E + .205 \ln GW + .084 \ln I$$
$$(-3.759)^{*} \quad (-1.438)^{*} \quad (.229) \quad (.746) \quad (.636)$$

$$+ .140 \ln NW + 1.159 \ln SY + .087 \ln D$$
$$(3.987)^{*} \quad (2.062)^{*} \quad (.776) \qquad R^2 = .565$$

3) $$\ln N/I = -2.00 - .678 \ln r(C|A) - .467 \ln S - .111 \ln E + .349 \ln GW + .275 \ln I$$
$$(-1.502)^{*} \quad (-2.076)^{*} \quad (-.310) \quad (1.132) \quad (2.028)^{*}$$

$$+ .121 \ln NW + .129 \ln SY + .208 \ln D$$
$$(3.006)^{*} \quad (1.944)^{*} \quad (1.590)^{*} \qquad R^2 = .440$$

4) $$\ln N/I = -1.39 - .365 \ln r(A) - .267 \ln r(C|A) - .285 \ln S + .088 \ln E + .200 \ln GW$$
$$(-3.316)^{*} \quad (-.634) \quad (-1.368)^{*} \quad (.272) \quad (.719)$$

$$+ .082 \ln I + .142 \ln NW + 1.222 \ln SY + .077 \ln D$$
$$(.616) \quad (3.887)^{*} \quad (1.886)^{*} \quad (.626) \qquad R^2 = .556$$

* Significant at the 10 percent level with a one-tailed test.

N = Total number of robberies, burglaries and larcenies over \$50 in 1968 reported to the FBI by local police.

A = Number of arrests in 1968 for the crimes of robbery, burglary and larceny over \$50 as reported to the FBI by local police departments.

C = Number of convictions in 1968 for the crimes of robbery, burglary, larceny over \$50 as reported by the local police to the FBI.

$r(A) = A/N$

$r(AC) = C/N$

$r(C|A) = C/A$

GW = Yearly income in thousands of dollars from the 1968 *County Business Patterns;* county wide.

E = Yearly sales per retail establishment in thousands of dollars taken from the 1967 *Census of Business.*

S = The average sentence served by inmates released from State and Federal Institutes who had been charged with either robbery, burglary, or larceny over \$50. This is state wide.

I = Community population, in thousands, taken from the estimates of 1968 population made by Rand McNally and published in their *Commercial Atlas and Marketing Guide.*

NW = The percent of the city population which was nonwhite in 1960, taken from the 1960 *Census of Population.*

SY = Mean school years completed by residents of the community in 1960 taken from the 1960 *Census of Population.*

D = Population density, persons per square mile, for the community taken from the 1960 *Census of Population.*

pected loss of utility is less than the utility of the expected loss, and that the cost of arrest is large, it follows that the change in the expected loss of utility will be greater for a 1 percent increase in $r(A)$ than for a 1 percent increase in the cost of punishment. This implies that more crimes will be prevented by increasing $r(A)$ than by increasing the cost of punishment or that the elasticity of N/I with respect to $r(A)$ should be greater than with respect to S, which is our measure of the cost of punishment.

The coefficient of E, retail sales per establishment, is positive in three of the four equations in Table 1. Retail sales per establishment is admittedly a very crude measure of the gain to criminal activity, a statement which is borne out by the low values of the t-statistic.

The only result of the analysis which ap-

pears to be in conflict with the theory is the coefficient for income, which we find to be positive, whereas a negative coefficient is anticipated. If it is true, as previously suggested, that *GW* is a measure of the gain to both legal and illegal activity, the coefficient of income could be either positive or negative.

To test our theory further, we ran regressions employing two additional variables, the labor force unemployment rate, *UE*, and the percent of families with incomes below $3,000, *PCBT*, for which we would expect positive coefficients. As seen in Table 2, the coefficients for *UE* are all positive while the coefficients for *PCBT*, however, are mixed.

Let the two roles that income plays in our model be designated as the "gain effect," i.e., as a measure of the gain from crime, and the "need effect," i.e., as a measure of the gain from legal activity. The results presented in Table 2 provide some evidence

TABLE 2—REGRESSION EQUATIONS USING *PCBT* AND *UE*[a]

(*t*-statistics in parentheses)

1) $$\ln N/I = -1.13 - \underset{(-3.601)^*}{.384\,\ln r(A)} - \underset{(-1.280)^*}{.268\,\ln S} + \underset{(.431)}{.137\,\ln E} + \underset{(.550)}{.175\,\ln GW} + \underset{(.634)}{.085\,\ln I}$$
$$+ \underset{(3.202)^*}{.150\,\ln NW} + \underset{(1.772)^*}{1.011\,\ln SY} + \underset{(.434)}{.050\,\ln D} - \underset{(-.121)}{.021\,\ln PCBT} \qquad R^2 = .561$$

2) $$\ln N/I = -1.15 - \underset{(-3.686)^*}{.338\,\ln r(A)} - \underset{(-1.231)}{.253\,\ln S} + \underset{(.707)}{.204\,\ln E} + \underset{(.622)}{.083\,\ln I} + \underset{(3.560)^*}{.158\,\ln NW}$$
$$+ \underset{(1.827)^*}{1.031\,\ln SY} + \underset{(.450)}{.052\,\ln D} - \underset{(-.439)}{.066\,\ln PCBT} \qquad R^2 = .558$$

3) $$\ln N/I = -1.33 - \underset{(-3.649)^*}{.354\,\ln r(A)} - \underset{(-1.046)}{.200\,\ln S} + \underset{(.357)}{.103\,\ln E} + \underset{(.938)}{.239\,\ln GW} + \underset{(1.452)^*}{.185\,\ln I}$$
$$+ \underset{(4.156)^*}{.137\,\ln NW} + \underset{(1.580)^*}{.826\,\ln SY} + \underset{(.015)}{.002\,\ln D} + \underset{(2.822)^*}{.343\,\ln UE} \qquad R^2 = .637$$

4) $$\ln N/I = -1.65 - \underset{(-3.857)^*}{.368\,\ln r(A)} - \underset{(-.981)}{.186\,\ln S} + \underset{(.850)}{.220\,\ln E} + \underset{(1.452)^*}{.184\,\ln I} + \underset{(4.161)^*}{.137\,\ln NW}$$
$$+ \underset{(1.644)^*}{.857\,\ln SY} + \underset{(.040)}{.004\,\ln D} + \underset{(2.772)^*}{.336\,\ln UE} \qquad R^2 = .629$$

5) $$\ln N/I = -1.47 - \underset{(-3.608)^*}{.356\,\ln r(A)} - \underset{(-1.051)}{.204\,\ln S} + \underset{(.347)}{.101\,\ln E} + \underset{(.927)}{.273\,\ln GW} + \underset{(1.453)^*}{.189\,\ln I}$$
$$+ \underset{(2.964)^*}{.130\,\ln NW} + \underset{(1.546)^*}{.820\,\ln SY} + \underset{(.006)}{.001\,\ln D} + \underset{(2.794)^*}{.347\,\ln UE} + \underset{(.240)}{.383\,\ln PCBT} \qquad R^2 = .638$$

6) $$\ln N/I = -1.49 - \underset{(-3.709)^*}{.364\,\ln r(A)} - \underset{(-.956)}{.184\,\ln S} + \underset{(.773)}{.207\,\ln E} + \underset{(1.398)^*}{.181\,\ln I} + \underset{(3.432)^*}{.142\,\ln NW}$$
$$+ \underset{(1.627)^*}{.858\,\ln SY} + \underset{(.043)}{.005\,\ln D} + \underset{(2.708)^*}{.334\,\ln UE} - \underset{(-.237)}{.033\,\ln PCBT} \qquad R^2 = .629$$

* Significant at 10 percent level with a one-tailed test.

[a] In addition to the variables defined in Table 1, Table 2 includes:

PCBT = Percent of the families in the community who had incomes below $3,000 in 1959 taken from the 1960 *Census of Population*.

UE = The labor force unemployment rate in 1968 generated from data published by the Bureau of Labor Statistics.

The sample size for both regressions is 53, as is reported in the text.

that the coefficient of income is picking up the gain effect. When *PCBT* and/or *UE* are used with *GW*, part of the need effect should be removed from the coefficient of *GW*, resulting in a larger coefficient for *GW* which we find when comparing equation 1) of Table 1 with the equations in Table 2. Likewise, when *GW* is removed from the equation, the coefficients for *UE* and *PCBT* fall, since they are now picking up the gain effect. This would also explain two of the negative coefficients for *PCBT*.

We also note that the coefficient for *E* moves in the opposite direction of the movement of the coefficient of *GW*, providing further evidence for our explanation of the positive coefficient for income.

IV. Conclusions

Following the lead of Becker and Fleisher, we have postulated an economic theory of property crime. Employing a Von Neumann-Morgenstern utility function we developed a model in which an individual must allocate his time between legal and illegal activities so as to maximize expected utility. We tested the model by running regressions on a cross-sectional sample of communities.

The empirical results give at least tentative credence to our hypothesis. We find that an increase in the probability of arrest and conviction and an increase in the cost of crime (punishment) both result in a decrease in the number of major property crimes committed.

Given our results, further work in this area does not appear to be out of order, particularly more sophisticated work employing better data.

REFERENCES

G. S. Becker, "Crime and Punishment: An Economic Approach," *J. Polit. Econ.*, Mar./Apr. 1963, *76*, 169–217.

B. Fleisher, *The Economics of Delinquency*, Chicago 1966.

E. Green, *Judicial Attitudes in Sentencing*, London 1961.

G. Schubert, *Judicial Behavior: A Reader in Theory and Research*, Chicago 1964.

T. Sellin and M. Wolfgang, *The Measurement of Delinquency*, New York 1964.

H. Theil, *Linear Aggregation of Economic Relations*, Amsterdam 1954.

J. Tobin, "A Statistical Demand Function for Food in the U.S.A.," *J. Royal Statist. Soc.*, Series A, Part II, 1950, *113*, 113–45.

Federal Bureau of Investigation, *Uniform Crime Reports—1968*, Washington 1969.

Federal Bureau of Prisons, *Inmates Released from State and Federal Institutions 1960*, Washington 1960.

National Opinion Research Center, *Criminal Victimization in the United States: A Report of a National Survey*, Washington 1967.

Rand McNally and Company, *1968 Commercial Atlas and Marketing Guide*, 99th ed. New York 1968.

U.S. Bureau of the Census, *Census of Business, 1967*, Vol. II, *Retail Trade, Area Statistics*, Washington 1970.

———, *County Business Patterns, 1968*, Washington 1969.

———, *U.S. Census of Population: 1960*, Vol. I, *Characteristics of the Population*, Washington 1963.

U.S. Bureau of Labor Statistics, *Employment and Earnings*, Apr. 1968, *14*.

A Skeptical Note on "The Optimality" of Wage Subsidy Programs

By IRWIN GARFINKEL*

In "Optimal Mechanisms for Income Transfers," a recent paper in this *Review*, Richard Zeckhauser purports to demonstrate the welfare superiority of wage subsidies to negative income tax programs. He assumes that taxpayers derive utility both from increases in the incomes and from increases in the hours worked of poor people. In his simple model of a single beneficiary and a single taxpayer, he shows that wage subsidy (*WS*) programs are less leisure inducing than equally costly negative income tax (*NIT*) programs, and are, therefore, Pareto superior.

There are at least three reasons for being skeptical. First, Zeckhauser and others[1] notwithstanding, there is no theoretical basis for concluding that pure *WS* programs involve less static work disincentive than pure *NIT* programs.[2] Second, Zeckhauser ignores the fact that the cost of *WS* and *NIT* programs depends not only upon the differential static work incentive effects of the two programs but as well upon the differential human investment incentive effects. Third, he also ignores the differential effects of demand on the costs of *WS* and *NIT* programs. In this note I show that: 1) the analytical confusion about the work incentive effects of *WS* and *NIT* programs arises from attempting to generalize from the single beneficiary case to the many beneficiaries case;[3] 2) the human investment incentive effects of *NIT* programs could possibly be superior to those of *WS* programs; and 3) the explicit introduction of demand into the analysis may also eliminate the alleged welfare superiority of *WS* programs.

Before proceeding with the analysis, it will be useful to describe algebraically the alternative economic opportunities which are available to individuals if there is no program, a *WS* program, or an *NIT* program. The budget constraints of individuals with income only from earnings, income from earnings plus an *NIT* grant, and income from earnings plus a *WS* grant are described algebraically below:

$$(1)\quad Y = HW$$

$$(2)\quad Y = G + (1 - m)HW \qquad \text{for } Y \leq \frac{G}{m}\text{; otherwise}$$
$$Y = HW,$$

$$(3)\quad Y = [g + (1 - n)W]H \qquad \text{for } W \leq \frac{g}{n}\text{; otherwise}$$
$$Y = HW,$$

* The research reported here was supported by funds granted to the Institute for Research on Poverty at the University of Wisconsin by the Office of Economic Opportunity pursuant to the provisions of the Economic Opportunity Act of 1964. I wish to acknowledge the helpful comments and suggestions of Michael Barth, Glen Cain, Edward N. Gramlich, Stanley Masters, Larry Orr, and Harold Watts. The conclusions are solely my responsibility.

[1] Jonathan Kesselman concludes, "the wage subsidy offers less static disincentive to work than any income subsidy (that is, *NIT*) plans yet developed" (p. 276). Michael Barth and David Greenberg recognize in a footnote that if labor supply curves are backward bending, the above conclusion may not hold, yet in the text of their paper they state their agreement with the above conclusion. Zeckhauser and Peter Schuck simply assert that a *WS* has superior incentive effects.

[2] Barth and Greenberg demonstrate that mixed *WS-NIT* systems in which individuals may receive subsidies simultaneously from both a *WS* and *NIT* program may induce greater labor supply reductions than pure *NIT* systems. I demonstrated in my discussion paper that the kind of self-categorizing mixed *WS-NIT* system advocated by Zeckhauser and Schuck may also induce greater labor supply reductions than pure *NIT* systems. Unlike the pure *WS* systems discussed in the text, either kind of mixed *WS-NIT* program could induce greater labor supply reductions than pure *NIT* systems even if no beneficiaries had backward bending labor supply curves.

[3] Richard Musgrave clarified a similar confusion in the analysis of the differential work incentive effects of a proportional and progressive income tax by distinguishing between the single taxpayer and many taxpayers cases.

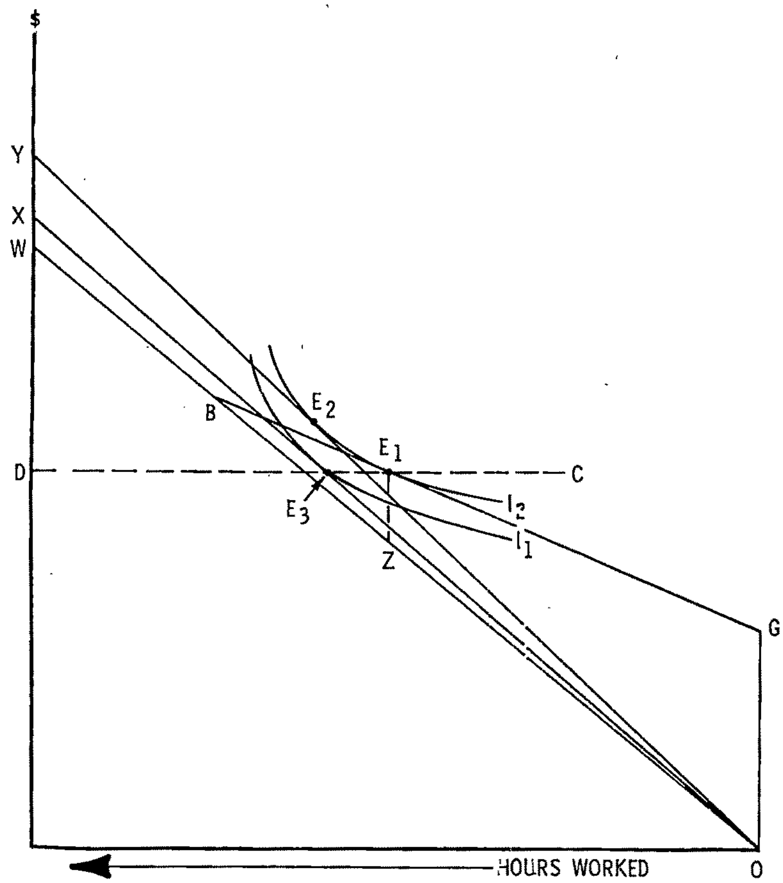


FIGURE 1. WORK INCENTIVE EFFECTS OF *WS* AND *NIT* PROGRAMS: SINGLE BENEFICIARY CASE

where Y = net income, H = hours worked, W = hourly wage rate, G = guarantee in the negative income tax (subsidy for a given family size when other income is equal to zero), m = marginal tax rate $(0<m<1)$ in the negative income tax program, g = the guaranteed wage rate (the minimum hourly wage rate the government will pay anyone irrespective of his market wage rate), and n = the marginal tax rate $(0<n<1)$ on wages in a wage subsidy. Individuals with earnings above G/m, the break-even level of income in the negative income tax program, receive no income subsidy. Individuals with wage rates above g/n, the break-even wage rate in the wage subsidy, receive no wage subsidy.

I. Work Incentive Effects

Single Beneficiary Case

It is easy to show given the conventional partial equilibrium assumptions that a single individual will work more under a *WS* than an *NIT* program if either his post-subsidy welfare, or his postsubsidy income, or his subsidy, is the same under the two programs. These three cases are demonstrated below with the aid of Figure 1.

Hours of work are measured from right to left along the horizontal axis and income along the vertical axis. It is assumed that initially all income is earned. The original budget constraint is given by the wage line *OW*. An *NIT* with a guarantee of *OG* dollars

results in a new budget constraint $OGBW$. Assume the new equilibrium lies along BG at a point such as E_1 where I_2 is tangent to GB. (The initial equilibrium is unspecified here except that it lies to the left of E_1 along OW.) Now consider a WS that would leave the individual indifferent between E_1 and a new equilibrium such as E_2. A WS pivots the old budget constraint or wage line OW, upwards from 0, resulting in a new wage line such as OY. The line OY must be steeper than GB since the WS increases the net wage while the NIT decreases the net wage. Consequently, given indifference curves with diminishing marginal rate of substitution, E_2 must lie to the left of E_1.

Now consider a WS that would leave the individual with the same income as the NIT program. Equilibrium must lie along the line DC, which passes through E_1 and which is parallel to the horizontal axis. Given indifference curves with diminishing marginal utility of leisure, only indifference curves which intersect DC to the left of E_1, will be steeper than the one at E_1. Since the new WS line must be steeper than GBW, equilibrium must lie to the left of E_1, at a point like E_3, where OX is tangent to I_1. The individual will work harder and be worse off under the wage subsidy.

Finally, it is now easy to see that if two WS and NIT programs are equally costly the former must lead to greater labor supply. The cost of the NIT program in the partial equilibrium framework of Figure 1 is equal to E_1Z, the net subsidy to the beneficiary. (The actual costs in general equilibrium will depend on the elasticity of the demand for labor. See Section III below.) Given indifference curves with diminishing marginal rates of substitution, equilibria on a WS budget line which entail a net subsidy as large as E_1Z must lie to the left of E_1.

Two (and Many) Beneficiaries Case

The major problem with the foregoing analysis is that while WS and NIT programs can be constructed so as to hold constant one individual's welfare or income, it is impossible to construct them in such a fashion that all individuals' welfare or income is held constant—unless wages and preferences of all individuals are identical. This point is extremely important, because it means that we cannot generalize from the single beneficiary case to the many beneficiaries case.

This is readily demonstrated by reference to Figure 2, which depicts the indifference curves of B_1, and of an additional potential WS or NIT beneficiary (B_2). While the wage rates of B_1 and B_2 are identical, their tastes differ. At low levels of income, B_2 has a much stronger preference for income vis-à-vis leisure than does B_1. Hence B_2's initial presubsidy equilibrium is at $E_0^{B_2}$ while B_1's initial equilibrium is at $E_0^{B_1}$.

The postsubsidy welfare of B_1 would be identical under the WS program which pivoted his budget constraint from OW to OY and the NIT program which shifted his budget constraint from OE to $OGBW$. But while B_2 would choose not to get any positive subsidy from the NIT program, he would necessarily benefit from the WS program. Thus while the NIT and WS programs hold B_1's postsubsidy welfare constant, they do not hold B_2's postsubsidy welfare constant. And, given differences in tastes and wage rates, unless special plans for every individual are devised, it is impossible to construct WS and NIT programs that would hold constant all individual potential beneficiaries' postsubsidy income or welfare.

Indeed Zeckhauser's simple model breaks down in the many beneficiaries case. For equally costly WS and NIT programs would lead in general to different distributions of welfare among potential beneficiaries. Without interpersonal utility comparisons, it is impossible to ascertain whether WS or NIT programs are Pareto superior.

More important perhaps, it is now easy to show that one cannot say a priori that WS programs must entail less static disincentive than NIT programs. As depicted in Figure 2, B_1 works more if the WS rather than the NIT program is enacted. But B_2 works less. Moreover as depicted, the magnitude of B_2's differential response, ab, exceeds, cd, that of B_1's. This constitutes a theoretical proof by counterexample that not all WS

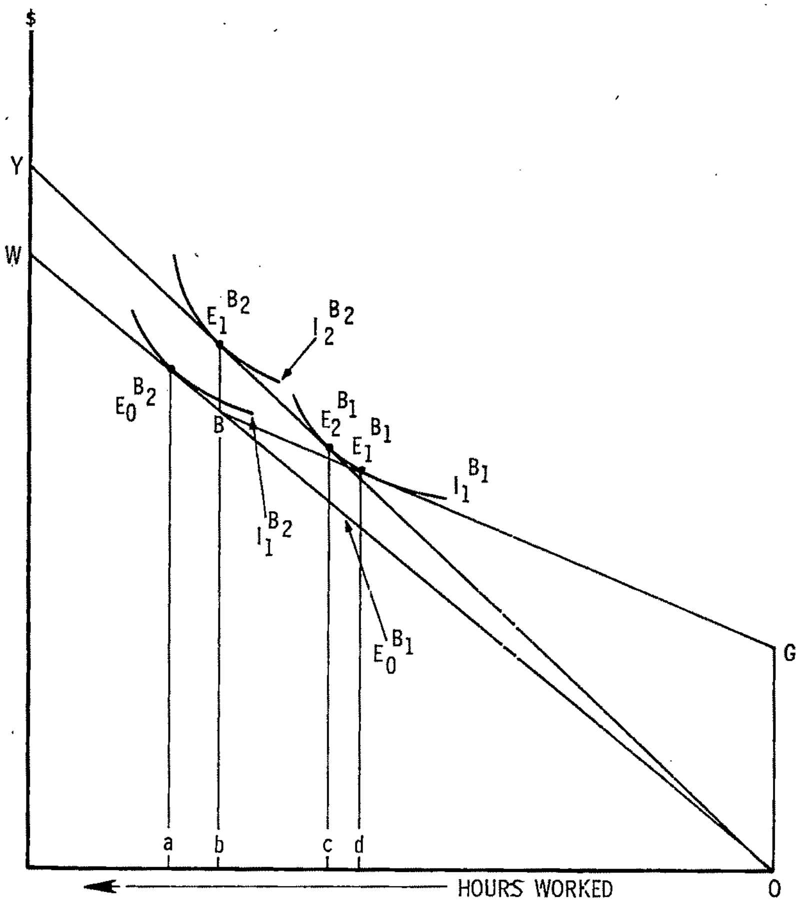


FIGURE 2. WORK INCENTIVE EFFECTS OF *WS* AND *NIT* PROGRAMS
TWO BENEFICIARIES CASE

programs entail less static disincentive than all *NIT* programs.

Note that as illustrated in Figure 2, the sum of the subsidies to B_1 and B_2 is greater in the *WS* than in the *NIT* program. Yet, it is easy to construct a plausible example in which the aggregate cost of the two programs is identical and the *WS* still entails more work disincentive. Imagine a third individual, B_3, whose wage rate and tastes are identical to B_2's, but whose work activity is restricted sufficiently by illness so that his subsidy from the *NIT* program would exceed his subsidy from the *WS* program by an amount which would equate the sum of the subsidies to B_1, B_2, and B_3 from the *WS* and *NIT* programs. Since B_3's labor supply would be determined by his health constraint and thereby is unaffected by either the *WS* or *NIT* program, we now have an example where an equally costly *WS* program entails more work disincentive than an *NIT* program.

My proof by counterexample depends upon a particular configuration of tastes, wage rates, and health status. The critical assumption is that some low wage individuals like B_2 have backward bending supply curves, yet refuse to participate in an *NIT* program because doing so would entail a reduction in their income.[4] This assumption

[4] If no individuals have backward bending supply curves, a pure *WS* system must necessarily entail less static work disincentive than a pure *NIT* system. For the *WS* will always be work inducing while the *NIT* will be leisure inducing provided leisure is a normal good.

and the others are, I believe, quite plausible. But the important point is that whether or not *WS* programs will entail less work disincentives than equally costly *NIT* programs is an empirical rather than a theoretical issue.

II. Human Investment Incentive Effects

The equilibrium condition for a utility maximizing individual who is considering a human investment decision is[5]

$$\frac{B}{(1+r)^t} = C \tag{4}$$

where B = benefit, C = cost, r = discount rate, and t = time period. That is, the individual will undertake the investment as long as the present value of the benefits exceeds the present value of the costs, where for simplicity all the costs are assumed to take place in the initial period t_0 and all of the benefits in period t_1. Let

$$B = H(W_T - W_0)_{t_1} + (NPB_T - NPB_0)_{t_1} \tag{5}$$

$$C = (H_0W_0 - H_TW_T)_{t_0} + D_{t_0} \tag{6}$$

where W = the wage rate, NPB = nonpecuniary benefits, H = hours worked, D = direct costs, and the subscripts T and 0 indicate respectively, with and without training. Note that in period t_0, H_T and W_T refer to hours and wage rates during training, while in t_1 they refer to hours and wage rates subsequent to training. Substituting (5) and (6) into (4) yields

$$\frac{H(W_T - W_0)_{t_1} + (NPB_T - NPB_0)_{t_1}}{(1+r)^t} = (H_0W_0 - H_TW_T)_{t_0} + D_{t_0} \tag{7}$$

On-the-job investment or training (*OJT*) is defined as a human investment which entails a reduction in W_0. Off-the-job investment is defined as an investment that entails a reduction in H_0.[6]

Enactment of either a *WS* or an *NIT* program would increase incentives to obtain *OJT*.[7] Consider first the effect of a *WS* program. If both the pretraining wage rate in the absence of training, W_0, and the posttraining wage rate, W_T, are below the break-even wage rate, and nonpecuniary benefits do not increase as a result of training, the *WS* program will reduce benefits by n percent. But enactment of the *WS* program would also reduce the opportunity costs, $H(W_0-W_T)_{t_0}$, of *OJT* by n percent. If we assume there are no direct costs to *OJT*,[8] therefore, the enactment of a *WS* program would not reduce the incentives of any workers to obtain *OJT*. Moreover, if either the pretraining wage rate in the absence of training or the posttraining wage rate are above the break-even wage rate, or if nonpecuniary benefits increase as a result of training, the *WS* program will reduce benefits by less than n percent and thereby increase the incentive to obtain *OJT*. Similarly enactment of an *NIT* will reduce the costs of *OJT* by m percent and the benefits by up to, but often by less than, m percent. Without knowledge about program parameters and individual preferences it is impossible to specify whether enactment of a *WS* or an *NIT* would lead to the greatest increase in incentives to obtain *OJT*.

A *WS*, however, would definitely provide greater *disincentives* than an *NIT* to obtain off-the-job human investment. The benefits of such an investment would again be reduced by up to n and m percent, respectively by enactment of either a *WS* or an *NIT* program. But the enactment of any *WS* would also increase the cost of off-the-job human investment by increasing the value of fore-

[5] I ignore the problems created by lumpiness of both costs and benefits.

[6] Although off-the-job investment need not entail reductions in hours worked, nearly all forms of such investment do involve reductions in hours.

[7] If there are no costs to training, the rational individual will always undertake the training so long as the benefits are greater than zero. The mere process of working may entail training. Whether this is true of the jobs held by poor workers is open to question. But if this were the case and if an *NIT* were more leisure inducing than a *WS* in the aggregate, it could be argued that more on-the-job training would take place if a *WS* rather than an *NIT* program were enacted because workers would be on the job more of the time in the former program.

[8] If the direct costs are greater than zero, the total costs would be reduced by less than n percent and there might actually be a disincentive.

gone earnings. In contrast, enactment of an *NIT* would never increase, and in many cases would reduce by up to *m* percent, the costs of off-the-job human investment. Thus, even if the enactment of an *NIT* program would discourage such investment because it reduced the costs of an investment by a smaller percentage than the benefits, the disincentives would be much smaller than those engendered by the enactment of a *WS* program.

Given the larger off-the-job human investment incentive effects of a *WS*, it is possible that a *WS* which appears to cost the same as an *NIT* when only their static work incentive effects are considered, may actually be more costly. Again, whether the overall human investment incentives effects of an *NIT* are superior to those of a *WS* is an empirical question which would depend upon such factors as program parameters, tastes, and the relative payoffs to poor workers of on- and off-the-job human investments.[9]

III. General Equilibrium Effects

Finally, it is impossible to analyze the benefits to beneficiaries or the costs to taxpayers of *WS* and *NIT* programs within the partial equilibrium framework employed by Zeckhauser.[10] If *WS* programs do lead to increases in the labor supply of poor people, market wage rates will fall, subsidy inclusive wage rates will fall, and the cost of the program will increase unless the demand for labor is perfectly elastic. The greater is the inelasticity of demand, the smaller are the benefits to the poor and the larger are the costs to the taxpayer of *WS* programs. Since on the other hand *NIT* programs must decrease the supply of labor, the more inelastic is the demand for labor, the larger are the increases in beneficiary wage rates and the smaller are the *NIT* program costs to the taxpayer. (Of course, *WS* programs could also lead to such results if in the aggregate like *NIT* programs, they turn out to be leisure inducing.) If the demand for labor is as inelastic as many studies suggest,[11] even with highly inelastic labor supply curves, very costly *WS* programs might lead to relatively small increases in the total incomes of the poor. The welfare superiority of *WS* to *NIT* programs in these circumstances is doubtful. In any case, it is clear that one cannot compare the welfare effects of *WS* and *NIT* programs without considering the elasticity of the demand for labor. The partial equilibrium costs and benefits of *WS* and *NIT* programs identified by Zeckhauser will only be equal to the general equilibrium costs and benefits if the demand for the labor of the poor is perfectly elastic.

I conclude by simply noting that since Zeckhauser's "demonstration" of the optimality of wage subsidy programs depends explicitly on the alleged superior work incentive effects of *WS* vis-à-vis *NIT* programs, and implicitly on the assumptions that the programs have no differential human investment incentive effects, and that the demand for labor is perfectly elastic, the case for the optimality of wage subsidy programs is open to serious question.

REFERENCES

M. Barth and D. Greenberg, "Incentive Effects of Some Pure and Mixed Transfer Systems," *J. Hum. Resources,* spring 1971, *6,* 149–70.

E. R. Berndt and L. R. Christensen, "Testing for the Existence of a Consistent Aggregate Index of Labor Inputs," mimeo, Social Systems Research Institute, Univ. Wisconsin, Oct. 1972.

[9] While the returns to *OJT* appear in general to be large, many poor workers are apparently poor precisely because their jobs have very little *OJT* component to them. On the other hand, the returns to poor workers from their current off-the-job human investments may also be small. It is clear that the differential human investment incentive effects on the poor of *WS* and *NIT* programs is a very complicated empirical question.

[10] I am indebted to Edward Gramlich and Michael Barth for calling my attention to the importance of the demand for labor in evaluating *WS* programs.

[11] See S. W. Black and H. H. Kelejian, especially fn. 5, p. 721, for a brief summary of estimates of the elasticity of demand for all labor. Their own estimate is .364. Two as yet unpublished studies by Edward Gramlich and Ernst Berndt and Laurits Christensen which attempt to estimate the elasticity of substitution for low-wage labor use different data sources and methodologies and derive very different estimates. Gramlich estimates that demand is highly inelastic while Berndt and Christensen estimate that demand is highly elastic.

S. W. Black and H. H. Kelejian, "A Macro Model of the U.S. Labor Market," *Econometrica*, Sept. 1970, *38*, 712–41.

I. Garfinkel, "The Work and Human Investment Incentives of Negative Income Tax and Wage Subsidy Programs," Institute for Research on Poverty disc. pap. 70–71, Univ. Wisconsin 1971.

E. N. Gramlich, "The Demand for Different Skill Classes of Labor," unpublished 1972.

J. Kesselman, "Labor Supply Effects of Income, Income-Work and Wage Subsidies," *J. Hum. Resources*, summer 1969, *4*, 275–92.

R. Musgrave, *The Theory of Public Finance*, New York 1959.

R. Zeckhauser, "Optimal Mechanisms for Income Transfers," *Amer. Econ. Rev.*, June 1971, *61*, 324–34.

——— and P. Schuck, "An Alternative to the Nixon Income Maintenance Plan," *Publ. Interest*, spring 1970, *19*, 120–30.

Optimal Mechanisms for Income Transfer: Note

By ROBERT E. SCHLENKER*

In a recent article in this *Review*, Richard Zeckhauser examines the subject of income transfer and concludes that a wage subsidy scheme (WS) is superior to a negative income tax plan (NIT). Zeckhauser's results are derived from a model in which a single representative poor man faces a single, fixed wage rate. The purpose of this comment is to extend Zeckhauser's discussion by comparing the two plans in a setting of many potential recipients, some unemployable and the rest facing different wages rates. When this is done, the superiority of the WS approach over the NIT does not appear quite as clear-cut as suggested in Zeckhauser's article.

The many-recipient setting requires two modifications in Zeckhauser's WS plan. First, a positive income floor is needed to cover those who are unemployable or who can work only a small number of hours. Under the scheme presented in Zeckhauser's article, with a negative basic stipend, such individuals would actually be taxed to subsidize the more affluent poor. Second, the subsidy must vary with the wage rate in such a way as to provide a smooth transition from subsidized and unsubsidized rates. The NIT deals with both those problems by providing an income floor and a break-even earnings level at which the subsidy ceases. Zeckhauser is also well aware of the problems and, with Peter Schuck, has suggested a realistic operational WS plan in another writing. The structure of that plan forms the basis for our subsequent discussion.

I. An Operational Wage Subsidy Plan

Zeckhauser and Schuck propose a two-plan approach as illustrative of a more general multiple-plan approach. One plan provides a relatively large basic stipend and places a high marginal tax rate on earnings; the other plan combines a small basic stipend with a wage subsidy (or negative tax rate on earnings). The subsidy varies with the wage rate so as to make up half the difference between the actual wage rate and \$3 per hour. The choice of plan would be left to each recipient. The two-plan approach thus deals with the situation of many recipients by providing an income floor and a wage rate subsidy which declines as the wage rate increases.

Assuming the poor are economically rational, those with low potential earnings would choose the plan with the high basic grant and high tax rate, while those with higher potential earnings would choose the wage subsidy option. Assuming, for simplicity, that the high stipend, high tax rate option incorporates a 100 percent tax rate, a client's income would then be represented by:

$$(1) \quad y = [w + b(w_0 - w)]h, \qquad \text{for } y > g, \quad \text{and} \quad w < w_0,$$

and

$$(2) \qquad y = g \text{ otherwise}$$

In these equations, y is the total income of the recipient g is the income floor (the basic stipend under the high stipend-high tax rate plan); w_0 is the break-even wage rate, or the rate at which the wage subsidy ceases; b is the fraction of the difference between the actual wage rate and the break-even wage rate which is made up by the subsidy; w is the actual wage rate; and h represents the actual number of hours worked. In Zeckhauser and Schuck, $b = .5$ and $w_0 = \$3$. Equation (1) is easily derived (see also the article by Jonathan Kesselman) from the fact that the subsidy, $y - wh$ (i.e., total income minus earned income), is designed to make up the

* Senior staff member of Inter Study, a policy research organization in Minneapolis, Minnesota. Helpful comments were received from a reviewer.

fraction, b, of the gap between w_0 and w. Therefore, $y-wh$ equals $b(w_0-w)h$ for $w<w_0$ and for all h, and solving for y yields equation (1).[1]

II. A Comparative Negative Income Tax Plan

For our purposes, the most useful way to compare the WS and NIT income transfer approaches is to provide them with identical boundary points on the range of earnings subsidized. That is, we postulate the same income floor and the same break-even point for each plan.

Because the break-even condition is defined in terms of earnings for the NIT and in terms of the wage rate for the WS, an additional specification is needed to link the two together. We do this by specifying h_0 as a reasonable maximum for the number of hours worked (such as 2,000 hours per year), and stating the condition, $w_0h_0=y_0$, where y_0 is the break-even earnings level under the NIT. If h_0 is a realistic maximum for h, then neither plan will subsidize earnings above y_0.

With these specifications, the equation for total income under the simplest form of a comparative NIT is:

$$(3) \qquad y = g + ewh$$

where g is again the income floor and e is a constant fraction determined by the income floor and break-even point as $(y_0-g)/y_0$. While the WS and the NIT plans we have specified will have identical income floors and identical break-even earnings levels, within the subsidized earnings range, the plans will be quite different.

[1] Equation (1) is noticeably different from the formulation given in Zeckhauser's (1971) model because of the necessity of dealing with different wage rates. Zeckhauser's WS equation is $y=s+hw(1-t)$, where s is the basic stipend and t is the marginal tax rate; both s and t are negative. In the multiple wage rate case, we must require that wage rates less than the break-even rate, w_0, although subsidized, result in effective wage rates which are less than w_0 and rise toward w_0 as w rises toward w_0. Therefore $w(1-t)$ must be positive, less than w_0, and rise toward w_0 as w rises toward w_0. Specifically, at $w=w_0$, t must equal 0. Also, at $w=w_0$, the income subsidy must vanish and $y=w_0h=s+hw_0(1-t)$. With $t=0$, this implies that $s=0$. With $s=0$ and $t=-b(w_0/w-1)$, with $0<b<1$, the desired conditions are met and equation (1) results.

III. Taxes and Subsidies Under the Two Plans

Under both plans, the recipient can increase his income in only two ways: increasing hours worked or the wage rate. The effects of an increase in hours worked will be examined first. From equations (1)–(3), partial differentiation of y with respect to h yields:

$$(4) \qquad WS\colon \frac{\partial y}{\partial h} = w + b(w_0 - w) \quad \text{for } y > g$$
$$= 0 \text{ otherwise}$$
$$NIT\colon \frac{\partial y}{\partial h} = ew$$

Since $w_0>w$ and $e<1$, $w+b(w_0-w)>w$ and $ew<w$. Thus, as expected, the NIT taxes increases in hours worked while the WS subsidizes such increases if $y>g$. However, if $y\leq g$, the WS plan levies a ***higher*** tax rate on increases in hours worked than does the NIT. The need for an income floor thus significantly affects the relative merits of the WS and NIT plans, and the role of the floor deserves further consideration.

The Role of the Income Floor

With an income floor, low wage employment would be discouraged under a WS plan, because a 100 percent tax rate would apply until after-transfer income rose above g. A lower tax rate would apply to the NIT. Thus, some persons could be expected to work under NIT who would not work under WS. For such persons, the NIT would yield higher total incomes and lower transfer costs than would the WS. (The transfer cost under NIT is $g-(1-e)wh$; under WS it is g for such persons.)

To encourage more employment at low wage rates, a high value of b could be utilized. The only restriction on b under WS is that it be between 0 and 1. To encourage employment, the after-transfer income from employment would have to be greater than the income floor by a significant amount even at low earnings levels. From equations (1) and (2), this would require that $[w+b(w_0-w)]h>g$ even for low wage rates and hours worked; the greater the value of b, the lower

the values of w and h at which the inequality would hold. How large a value of b is required? This is an empirical question on which we have no information but the examples in Table 1 provide some insights.

The table indicates the number of hours of work required before the income from work with a wage subsidy exceeds the value of the income floor. The hours are calculated for different wage rates, different values of b, and different income floors. The general point is that a relatively high value of b may be necessary in order to encourage employment at low wage rates. For example, if 2,000 hours represents full-time work (on an annual basis) and \$3,000 is the floor, at least half-time work is required under all values of b before work income exceeds the floor for wage rates less than \$2. If the floor is \$2,000 the number of hours required is smaller. In either case, the greater the value of b, the lower the required number of hours.

Table 1—Minimum Hours of Work Necessary Under *WS* for Income to Exceed Floor

	$b=.25$	$b=.50$	$b=.75$	$b=.90$
Case 1. $g=3,000$, $y_0=6,000$, $w_0=3$, $e=.5$				
$w=1$	2000	1500	1200	1071
2	1333	1200	1091	1034
Case 2. $g=2,000$, $y_0=6,000$, $w_0=3$, $e=.67$				
$w=1$	1333	1000	800	714
2	889	800	727	690

A high value of b may thus be required if low-wage employment is to be encouraged under a *WS* plan. There is a tradeoff involved, however, between increasing b and increasing the incentive for recipients to obtain wage rate increases, as discussed below.

The Wage Rate Tax

The second way the recipient can increase his income is by obtaining a wage rate increase. While major concern is usually with hours worked (see Zeckhauser), the tax on wage rate increases is not a trivial issue. For instance, under a *WS* plan, at a given wage rate, the transfer cost for a recipient can only be reduced if the recipient reduces his hours of work. The current welfare reform debate suggests that such action would be frowned upon by the nonpoor (even if program costs were thereby reduced). The only acceptable way to reduce costs would then be via wage rate increases. The emphasis on manpower programs in welfare reform proposals indicates the public attractiveness of programs which promise to increase the productivity of the poor. At the same time, the notable lack of success demonstrated by such programs in the past indicates that obtaining wage rate increases is not costless for either recipient or society. Taxing wage rate increases, by adversely affecting recipient incentives to obtain such increases, can only intensify the problem.

Both *WS* and *NIT* place a tax on wage rate increases. The relevant effective wage rates were given in equation (4) above. The change in the effective wage rate resulting from a change in the actual wage rate can be found by differentiation of (4) with respect to w:

$$(5)\qquad WS:\ \frac{d}{dw}\left(\frac{\partial y}{\partial h}\right) = 1 - b \quad \text{for } y > g$$
$$= 0 \quad \text{otherwise}$$
$$NIT: \qquad = e$$

The tax rate on wage rate increases is 1 minus the effective wage rate increases given in (5). Thus, the tax rate on wage rate increases is b under *WS* and $1-e$ under *NIT*. Under *WS* there is therefore a tradeoff between the size of the wage rate subsidy and the tax on wage rate increases. The greater b, the greater are both the subsidy and the tax.

In addition, when the *WS* and *NIT* are compared in terms of the tax on wage rate increases, it is evident that the value of b may have to be greater than $1-e$ in order to encourage work at the lower end of the earnings range.[2] This is illustrated in Table 1.

[2] Graphically, the relation between b and $1-e$ can be shown as follows. On a diagram with h measured along the horizontal axis, equation (1) of the *WS* plan would be a straight line beginning at the origin and rising with slope $w+b(w_0-w)$. The *NIT* plan line would begin at height g on the vertical axis and rise with a lower slope of ew. Since $e=(y_0-g)/y_0$ and $h_0w_0=y_0$, the two lines

In Case 1, b would be greater than $1-e$ if b were greater than .5; in Case 2, this would occur if b were greater than .33. In other words, an operational *WS* plan may result in a higher tax rate on wage rate increases than would be present under a comparable *NIT* plan.

Whether or not a higher *WS* wage rate tax means a recipient will be less likely under *WS* than *NIT* to seek a given wage rate increase depends, of course, not on the increase in the effective wage rate but on the increase in income. If the *WS* leads to a greater equilibrium number of hours worked at all wage rates, then the *income* increase resulting from a given wage rate increase might still be greater under *WS* than under *NIT*. The size of the incentive to increase wage rates under the two plans, the costs of obtaining the increases, and the difference in response under the two plans are obviously subjects for empirical investigation. The negative income tax experiments currently underway will, unfortunately, provide answers only for the *NIT* approach.

would intersect at

$$h = [y_0(1 - e)]/[w(1 - e - b) + bw_0]$$

where we have replaced g by $y_0(1-e)$. It can then be shown that for $w<w_0$ the intersection will occur at $h=h_0$ only if $b=1-e$. Similarly, for $w<w_0$ the intersection will occur at $h<h_0$ only if $b>1-e$. Comparisons of the two plans at the same number of hours worked are, of course, inappropriate, since a major objective of the *WS* is to lead to a greater number of hours worked than the *NIT*. The point of this calculation is only to emphasize the fact that if the *WS* is to lead to significantly greater income possibilities than the *NIT*, it may be necessary to levy a greater tax on wage rate increases under the *WS*.

IV. Conclusions

These remarks are not intended as an argument against the wage subsidy concept, for the idea has considerable merit. Rather, my major thrust is that a complete appraisal of alternative income transfer mechanisms must include considerations in addition to those raised by Zeckhauser. To reiterate the main points: an operational wage subsidy plan requires an income floor and a wage rate subsidy which declines as the wage rate increases. The existence of a floor may require high subsidies of low wage rates in order to encourage any work at these levels; the negative income tax may thus lead to more work by those facing low wage rates than would the wage subsidy. High subsidy rates at low wage rates, in turn, may result in a higher tax rate on wage rate increases than would exist under a negative income tax plan with the same income floor and break-even earnings level. Finally, the magnitude of and response to these various incentives are sufficiently unclear as to call for extensive empirical investigation.

REFERENCES

J. Kesselman, "Labor-Supply Effects of Income, Income-Work, and Wage Subsidies," *J. Hum. Resources*, summer 1969, *4*, 275–92.

R. Zeckhauser, "Optimal Mechanisms for Income Transfer," *Amer. Econ. Rev.*, June 1971, *61*, 324–34.

——— and P. Schuck, "An Alternative to the Nixon Income Maintenance Plan A Solution to the Problem of Work Incentives," *Publ. Interest*, spring 1970, *19*, 120–30.

On Terms of Foreign Borrowing

By PRANAB K. BARDHAN*

The theoretical literature on terms of foreign borrowing is rather narrow in the range of issues analyzed. Much of the recent literature is on an elaboration of the standard "optimum tariff" argument to the case where services of capital are internationally purchased.[1] It is now well-known that a borrowing country, if it is an important borrower in the international capital market, may gain by restricting its international borrowing; depending on the relevant elasticities, one can easily work out the optimum terms (or the interest rate) at which borrowing should be done so that the monopoly power (strictly, monopsony power in the purchase of capital services) in the international capital market is fully utilized.

For many borrowing countries, particularly in the underdeveloped world, the relevance of this analysis is, however, limited. These countries, taken individually, are often small borrowers in the international capital market and the task of setting on the national level the optimum terms of borrowing on the basis of their monopsony or oligopsony power is not particularly relevant. But a more significant limitation is that the whole analysis is static and ignores important time dimensions involved in problems of capital borrowing. The present paper concentrates on one such dimension.

The terms of a foreign loan involve not merely the interest rate but also a maturity period by which time the loan is to be paid back. In the real world a borrowing country is often confronted with a choice among alternative loan packages with varying rates of interest and lengths of maturity period. From the point of view of the long-run benefits of the borrowing country, how should one choose between a loan with, say, 5 percent rate of interest and a maturity period of 20 years and another loan with a higher, say, 7 percent rate of interest but a longer maturity period of, say, 30 years? Or, to put it in other words, in the long run is a rise in the interest rate on the foreign loan by a certain percentage costlier than a given shortening of the maturity period? This is an important practical problem in loan negotiations, and the existing theoretical literature on foreign borrowing does not throw much light on it.[2] The present paper tries to answer the problem in terms of a very simple dynamic model.

I take a Harrod-Domar growth model with its drastically simplifying assumptions of a constant capital-output ratio and the ratio of savings to national income. There is only one commodity, and problems of trade with incomplete specialization are ignored. The foreign loan under consideration is a once-for-all addition to the capital stock of the country. For simplification again, all capital is assumed to last forever, but the foreign loan has to be amortized in equal annual installments. I also ignore the gestation lag between availability of capital and its yield of output.

National income, Y, is given by

$$(1) \qquad Y = a(K + L) - iL\left(1 - \frac{t}{T}\right)$$

where a is the output-capital ratio, K is the domestically owned capital stock, L is the given amount of foreign loan, T is the maturity period of the loan, t is time, and i is the interest rate. The second term on the right-hand side of (1) is the drain on national income due to interest payments on the foreign loan.

* Professor of economics, Indian Statistical Institute, New Delhi. In formulation as well as analysis of the problem in this paper, I have benefitted from discussions with G. Arsenis and V. K. Sastry. All errors are, of course, mine alone.

[1] See, for example, Murray Kemp, ch. 13, and references cited there.

[2] The only two references in the theoretical literature where I have seen a brief mention of the problem are Abdul Quayum, p. 369, and Goran Ohlin, pp. 49–50.

Assuming that savings form a constant fraction, s, of national income and are instantaneously invested, over the loan period capital accumulation is given by

$$\frac{dK}{dt} = sY - \frac{L}{T} \tag{2}$$

where the second term on the right-hand side of (2) is the annual amortization payment. Integrating the differential equation (2), with the help of (1), and substituting the resulting value of K in Y, we rewrite (1) as

$$Y(\mathrm{t}) = \alpha e^{ast} + \beta \tag{3}$$

where $\alpha = aK(0) + (L/asT)(a-i)(asT-1)$
and $\beta = (L/asT)(a-i)$

If we now represent the index of long-run benefits by the total stream of national income over the length of the loan maturity period T, this is given by

$$y = \int_0^T Y(\mathrm{t})d\mathrm{t} \tag{4}$$

Using (3) in (4) and after integration,

$$y = \frac{\alpha(e^{asT} - 1)}{as} + \beta T \tag{5}$$

If we now want to find out the impact of a change in the interest rate charged on the foreign loan on this index of long-run benefits, we take[3]

$$y_i = \frac{\partial y}{\partial i} = \frac{L}{a^2 s^2 T} \cdot \left[(e^{asT} - 1)(1 - asT) - asT\right] < 0 \tag{6}$$

This means that a rise in the interest rate will reduce the long-run benefits from the loan, which is an expected result.

Similarly, if we want to find out the impact of a change in the length of the maturity period, T, on y, we take

$$y_T = \frac{\partial y}{\partial T} = \frac{L(a - i)(e^{asT} - 1 - asT)}{(asT)^2} + Y(T) \tag{7}$$

[3] Suppose $F(x) = (e^x - 1)(1 - x) - x$, then $F(0) = 0$ and $F'(x) < 0$ for $x > 0$; substituting $x = asT$, we can see how equation (6) is negative.

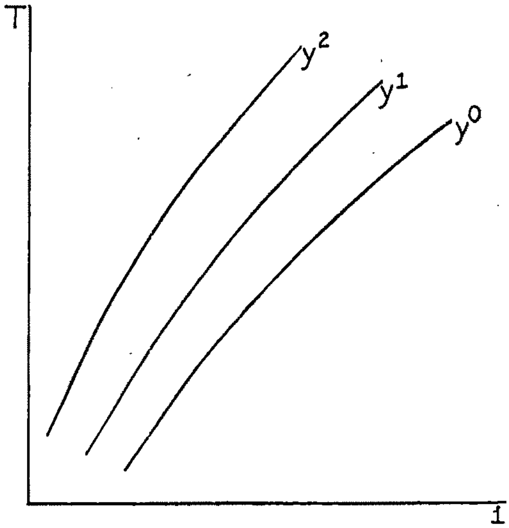


FIGURE 1. INDIFFERENCE CURVES BETWEEN INTEREST RATE AND MATURITY PERIOD

Equation (7) is positive if $a > i$, i.e., if the marginal product of capital is higher than the rate of interest, the longer the maturity period the better. In the opposite case it would be better to repay the loan as quickly as possible, in fact, it does not pay to borrow at all. We shall, therefore, assume $a > i$, and hence (7) is positive.

From (5) we can now work out a relationship between T and i for the same y; since y_i is negative and y_T positive, dT/di is positive. This means that if the total sum of income over the loan period is to be maintained, a higher interest rate on the foreign loan is to be matched by a longer maturity period. Figure 1 shows what may be called the indifference curves between T and i for different values of y. The more to the left a curve on this indifference map is, the higher is the level of the present value of national income associated with it. Thus, with the help of this indifference map one can now plot the coordinates of i and T for alternative loan packages and choose the combination that happens to be on the highest indifference curve.

One can also work out the value of the elasticity of substitution between i and T along such an indifference curve, as given by

$$\epsilon = \frac{dT}{di}\frac{i}{T} = \frac{-y_i i}{y_T T} \tag{8}$$

From (6) and (7) it is not difficult to show[4] that the value of ϵ is less than unity under the sufficient condition that the initial amortization payments do not exceed savings causing decumulation of capital, i.e., $dK(0)/dt \geq 0$, which we assume. This means that the offer of a loan with a maturity period 1 percent shorter than that in another loan must compensate with a *more* than 1 percent lower interest rate compared to the other to be equally attractive for the borrowing country. If there is offer of a loan with 5 percent interest rate and 20 years of maturity, another loan with 30 years of maturity, to be equally attractive from the point of view of long-run benefits of the borrowing country, may charge more than 7.5 percent as interest rate.

Another way of looking at the same result is that the proportionate effect of a rise in the interest rate on the index of long-run benefits is *less* than that of a shortening of the maturity period. This seems to be consistent with the prevailing feeling in international aid negotiations that the recipient countries tend to be more sensitive to changes in the maturity period of loans than to changes in the interest rate.

One objection to the index of long-run benefits we have chosen is that the time horizon exactly coincides with the maturity period. The general case where the long run exceeds the loan period is quite complicated. In one special case, however, all of our qualitative results above remain unchanged. This is the case where we rewrite equation (4) as

$$y = \int_0^T Y(t)dt + K(T) \tag{9}$$

The inclusion of the terminal capital stock in the index of long-run benefits is one way of representing the country's concern for income and consumption in the post-loan period. Since it is easy to check that the value of $K(T)$, which we obtain after integration in equation (2), is decreasing in i and increasing in T, and since $[K_i(T) + K_T(T)]$ is positive under the assumptions we have made so far, it is quite simple to show that, as before, $y_i<0$, $y_T>0$, $\epsilon<1$.

Another point about our index of long-run benefits is that it has been measured in terms of the simple sum of national income over the length of the loan maturity period. Implicitly we have assumed the social discount rate to be zero. What if the social discount rate is positive and the index of long-run benefits is given by the present value of the stream of national income over the loan period discounted at a positive constant rate ρ? Equation (4) is now replaced by

$$y = \int_0^T Y(t)e^{-\rho t}dt \tag{10}$$

and equation (5) by[5]

$$y = \frac{\alpha[e^{(as-\rho)T} - 1]}{as - \rho} + \frac{\beta(1 - e^{-\rho T})}{\rho} \tag{11}$$

If we now want to find out the impact of a change in the interest rate charged on the

[4] From (6) and (7),

$$y_T T + y_i i = Y(T)T + \frac{L}{as^2T}[(e^{asT} - 1 - asT) - isT(e^{asT} - 1)]$$

Putting the value of $Y(T)$ from (3), this is positive if

$$K(0) > L\left(\frac{1}{2 + asT} - \frac{(a-i)}{a}\right)$$

which is satisfied if $dK(0)/dt$ in equation (2) is nonnegative, i.e., if the initial amortization payment does not exceed savings causing decumulation of capital in the initial period. This, we may assume to be valid without much strain on the realism of the model.

[5] Equation (11) is derived under the assumption of $as \neq \rho$. When $as = \rho$,

$$y = \alpha T + \frac{\beta(1 - e^{-\rho T})}{\rho}$$

$$y_i = \frac{L}{\rho^2 T}[(1 - \rho T)\rho T - (1 - e^{-\rho T})] < 0$$

and

$$y_T = \frac{L(a - i)}{\rho^2 T^2}(\rho T - 1 + e^{-\rho T}) + e^{-\rho T}Y(T) > 0 \quad \text{for } a > i$$

foreign loan on the new index of long-run benefits, we take

$$(12)\quad y_i = \frac{Le^{-\rho T}}{asT}\left[\frac{(e^{asT} - e^{\rho T})(1 - asT)}{as - \rho} - \frac{(e^{\rho T} - 1)}{\rho}\right] < 0$$

Thus, the earlier result about the effect of a change in the interest rate goes through. As for the effect of a change in the length of the maturity period,

$$(13)\quad y_T = \frac{e^{-\rho T}L(a - i)}{asT^2}\left[\frac{(e^{asT} - e^{\rho T})}{as - \rho} - \frac{(e^{\rho T} - 1)}{\rho}\right] + e^{-\rho T}Y(T)$$

It is easy to show that (13) is positive under our assumption of $a>i$, again confirming our earlier result. This means that our earlier results about the indifference map in Figure 1 remain intact.

What about the elasticity of substitution, ϵ, between i and T? Here, under the following sufficient conditions[6] our earlier result of $\epsilon<1$ goes through:

(a) $\frac{dK}{dt}(0) \geq 0,$ which we have assumed before,

and

(b) $a \geq i\rho T$

Condition (b) is restrictive, but in most practical cases likely to be satisfied. For example, if the capital-output ratio is 3:1, the interest rate is 8 percent and the social discount rate is 10 percent, even as long a maturity period as 40 years will satisfy this sufficient condition. Under these sufficient conditions (a) and (b), the proportionate effect of a rise in the interest rate on the present value of national income is *less* than that of a shortening of the maturity period. Our earlier case $\rho=0$ is only a special case satisfying condition (b).

REFERENCES

M. C. Kemp, *The Pure Theory of International Trade and Investment*, Englewood Cliffs 1969.

G. Ohlin, *Aid and Indebtedness*, Paris 1966.

A. Qayum, "Long-Term Economic Criteria for Foreign Loans," *Econ. J.*, June 1966, *76*, 358–69.

[6] From (12) and (13), and after simplification,

$$y_T T + y_i i = \frac{Le^{-\rho T}}{asT}\left[\frac{(e^{asT} - e^{\rho T})(1 - isT)}{as - \rho} - \frac{(e^{\rho T} - 1)}{\rho} + \frac{sT^2}{L}Y(T)\right]$$

It is easy to show that this is positive if: (a) $(dK(0)/dt) \geq 0$ and (b) $a \geq i\rho T$, since from (2) and (3),

$$Y(T) = \frac{dK}{dT}(0)\frac{e^{asT}}{s} + \frac{L}{asT}[a + i(e^{asT} - 1)]$$

Recent Exercises in Growth Accounting: New Understanding or Dead End?

By RICHARD R. NELSON*

The growth accounting literature has been enriched recently by several major quantitative studies and a sophisticated technical dialogue.[1] The obvious high quality of this work may lead some economists to think that great progress has been made in our understanding of economic growth and that we are nearly home. I suggest that while recent research has increased our knowledge, studies of this sort have run into sharply diminishing returns and soon will arrive at a dead end leaving many essential open questions. The basic limitations of growth accounting discussed here are well known to many practitioners. One is the fundamental arbitrariness of the techniques that have been employed to distinguish between movements along a production function and shifts in that function. The second is the possible major misspecification involved in implicitly treating experienced growth as the sum of the contributions made by separate factors. These basic limitations have been repressed in the recent discussions and it seems important to bring them clearly into view again. Because of these limitations, growth accounting has not really illuminated some of the central questions of growth theory and growth policy, which I will discuss in the concluding section.

I. The Problem of Competing Explanations

From its beginnings, growth accounting has been concerned with trying to estimate how much of growth can be explained by movements along a production function, and how much should be attributed to advances in technological and organizational competence. However the early studies recognized quite explicitly the difficulties, perhaps even the theoretical impossibility, of distinguishing between alternative explanations of observed growth patterns without rather strong a priori assumptions.[2] Some of the recent studies appear to give the impression that on the basis of rather weak a priori assumptions there is a theoretically correct way of distinguishing movements along a production function from shifts in it. It seems important, therefore, to review the basic problem.

The difficulty can be seen sharply if one assumes the following stylized aggregative facts.[3] Output (*GNP*) has been growing at the same rate as capital and at a faster rate than labor; hence the capital-output ratio has been constant and output per worker and the capital-labor ratio have been rising. Factor shares have remained constant; thus the rate of return on capital has been constant and the wage rate has risen. These "facts" very roughly characterize the *U.S.* growth experience that the accounting exercises seek to explain. Consider the following two competing explanations, both consistent with the time-series data. One is that the underlying production function is Cobb-Douglas (of unitary elasticity of substitution) and technical change has been neutral in the sense of Hicks. The second is that the underlying production function has an elasticity of substitution less than one, and that technical change has been labor saving. The

* Professor of economics, Yale University. I am indebted to Carlos Diaz Alejandro, Robert Evenson, William Fellner, Yoav Kislev, and William Nordhaus for useful discussion and criticism. None of these necessarily agrees with all or any of the thrust of this paper.

[1] In particular there is the important study of Zvi Griliches and Dale Jorgenson (1967), discussion of this work by Edward Denison, and the attempt at reconciliation by Griliches and Jorgenson (1970). M. Ishaq Nadiri recently has presented a general discussion of the growth accounting literature.

[2] See for example the cautious remarks of John Kendrick

[3] Robert Solow (1970) makes use of these same stylized facts.

first interpretation is depicted in Figure 1, the second in Figure 2. Points (*a*) and (*b*) in the two figures are identical and the slopes of the curves (the marginal productivity of capital) at those points also are identical. Thus both interpretations are consistent with the input, output, and factor price data.

The two interpretations are different in the following "growth accounting" sense. In the case of Figure 1, output per worker would have grown by Δ_{11} if capital per worker had grown as it did, but the production function had not shifted. Δ_{12} represents the increase in output per worker not explained by growth of the capital-labor ratio and hence due, in some sense, to technical change. In Figure 2, Δ_{21} can be attributed to growth of capital per worker and Δ_{22} to technological change in the sense above. In the latter interpretation, the lower elasticity of substitution means that less of productivity growth can be attributed to growing capital intensity, hence more must be attributed to improved technology. As Peter Diamond, Daniel McFadden, and Miguel Rodriguez (among others) have pointed out, since both interpretations are equally consistent with the time-series data there is no way to choose among them, without a priori assumptions.

In fact, the growth accounting exercises have not proceeded by attempting to specify a particular "production function" and estimate its parameters. Rather the strategy is somehow to build up an input "index" that

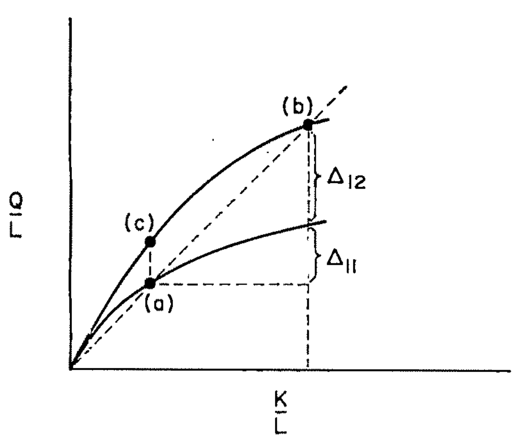


FIGURE 1

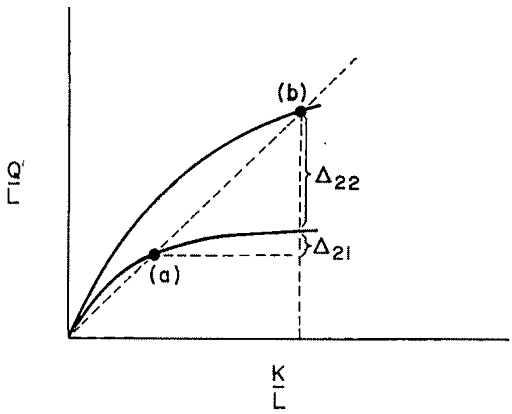


FIGURE 2

ALTERNATIVE INTERPRETATIONS OF PRODUCTIVITY GROWTH

measures the contribution of growth of input quantity (disaggregated so that "quality" changes can be accounted) to output growth without explicit commitment to a particular production function. There is a semantic problem here. The "*input* index" really is an estimate of *output* under the assumption of constant technology. In any case, this research methodology does not avoid the problem but simply evades it.

The use of a particular weighting or index scheme for input growth is the growth accountant's de facto assumption about the shape of the production function. A starting place for all growth accounting is the assumption that, if the neoclassical theory holds, at any time factor prices equal marginal productivities. Thus it is natural to weight factor inputs by their prices. But prices when? They are likely to vary over the period in question. There seems to be a belief that use of the Divisia index is the correct answer.[4] The Divisia index weights inputs at any moment of time by their prices at that moment. Ideally in using the Divisia method one would estimate the instantaneous growth of output due to factor increase by the method proposed by Solow (1957). Let

[4] Griliches and Jorgenson seem to claim this. The original statement of the correctness of the Divisia formula seems to be Marcel Richter's.

dQ^*/Q be instantaneous output growth due to input growth.

$$(1) \quad \frac{dQ^*}{Q^*} = S_L \frac{dL}{L} + S_K \frac{dK}{K}, \text{ where of course,}$$

$$\frac{dQ}{Q} = \frac{dQ^*}{Q^*} + \frac{dA}{A}$$

Integration of the right-hand side of equation (1) over the total time period would appear to give total output increase due to input growth (Δ_{11} of Figures 1 and 2).

$$(2) \quad \log Q^*(\mathrm{T}) - \log Q^*(0) =$$

$$\int_0^{\mathrm{T}} \left(S_L(\mathrm{t}) \frac{d \log L(\mathrm{t})}{dt} + S_K(\mathrm{t}) \frac{d \log k(\mathrm{t})}{dt} \right) dt$$

I shall argue that appearances are deceiving.

In practice factor price weights (shares) cannot be reestimated continuously but on a yearly (or other periodic) basis. If periods are short enough there is no real problem introduced here. The real problem is that we are not concerned with growth over a "year" but rather over a rather lengthy period, say the half century since the end of World War I. When considering a period of this length the fact that the overall period is divided up into short periods does not help; a large number of subperiods need to be added up over the total period. The problem does not go away. It is the finiteness of the total comparison period that causes the difficulty.

The time path of factor prices or shares over finite time is what it is because of both changes in factor ratios and technological change. The Divisia formula fails to distinguish between alternative explanations of factor prices. Thus under the interpretation of Figure 2 the capital share would have fallen but for the fact that technical change was capital using. If one wants to attribute to factor growth only what output growth would have been had technology been constant, then one must use in equation (2) not actual "shares," but the time path of shares as they would have been had technology not changed.[5] But to do this requires that one be able to specify the original production function which was the original impasse.

The route out of the impasse requires more specification based on other data. For example, one could attempt to estimate the elasticity of substitution from engineering design data. Or, if one had access to cross-section data on outputs and inputs as well as time-series data, under certain assumptions one might be able to sort out the shape of the production function from shifts in that function. Assumptions about momentum or independence can facilitate discrimination. For example, one could assume that the rate of technical change were constant. Then if that assumption were correct, and there were considerable variation over time in the rate of growth of the capital-labor ratio, one might be able to sort out among competing hypotheses. Fellner takes such an approach. In any case, in order to do growth accounting in a nonarbitrary way, we need knowledge that goes beyond the data that are used in the growth accounting.

II. The Meaning of Growth Attribution

Let us assume that the problem described above is solved. Then it would be possible to pose the following question: how much growth would we have experienced had only technology changed, or only capital, or only labor, or only capital and labor, or various other combinations of factors. However, it is uncertain what meaning one would give to the answers. If by attribution to a particular factor we mean the growth that would have occurred had it alone changed, the relative attribution to growth of different factors is not independent of the time period in ques-

[5] That is the shares need to be written explicitly as a function

$$S_i\left(\frac{K}{L}(\mathrm{t}), A(\mathrm{t})\right)$$

Equation (2) needs to be specified with

$$S_i\left(\frac{K}{L}(\mathrm{t}), A(0)\right)$$

The Richter specification of equation (2) does not do this.

tion, even if all factors are changing at a constant rate. Further, while using this meaning of attribution the sum of the attributions adds up to total growth for very small time periods, they may not add up to total growth if finite time periods are considered. These points were raised earlier by Herbert Levine and Benton Massell but seem to have been ignored in the recent discussion.

Assume that interpretation 1 is known to be correct, that the production function is Cobb-Douglas and technical advance is neutral, and that all factors were growing at constant rates. Assume that by attribution to a factor we mean the amount of output growth that would have occurred had that factor alone changed. Measure the instantaneous growth rates associated with the yearly growth rates of capital and labor. Label these λ_K and λ_L. Estimate the instantaneous rate of technical progress by the Solow method using the instantaneous rate of output growth. Call this λ_A. In the case of infinitesimally small changes the attribution of any factor of growth relative to another, or relative to total growth, is easily computed from these constants. All is nice and neat, with attributions to individual factors adding up to total growth.

All the neatness goes when finite time periods (say between 0 and T) are considered. For capital and technological change the attribution ratio is as follows:

$$(3)\quad \frac{\text{Attribution to Technical Advance}}{\text{Attribution to Capital Growth}} = \frac{e^{\lambda_A T} - 1}{e^{\lambda_K s_K T} - 1}$$

Relative attributions to different factors are sensitive to the time period in question. In the Cobb-Douglas case the percentage of total growth explained by growth of any particular factor, with the others held constant, shrinks to zero as the time period increases. For technical advance, for example:

$$(4)\quad \frac{\text{Attribution to Technical Advance}}{\text{Total Growth}} = \frac{e^{\lambda_A T} - 1}{e^{(\lambda_K s_K + \lambda_L s_L + \lambda_A)T} - 1}$$

This expression clearly goes to zero as T increases.

At first glance all this seems strange, but it is not. The problem is the same one that plagued the profession many years ago when it was trying to attribute total product (rather than growth) between the different factors. We learned then that this was impossible. We could attribute at the margin. But there was no way of attributing shares of the total.

To see the problem from another perspective look again at Figure 1. Δ_{11} measures how much output would have grown had technology remained constant. It is analogous to the numerator of equation (3). Δ_{12} is the measure of the contribution of technological change measured as a residual, but it does not measure how much output would have grown had capital remained constant. It is not analogous to the denominator in equation (3). The distance between points (*c*) and (*a*) measures that. Call this Δ_{12*}. This is the same as Δ_{12} only for very small changes. For finite changes, Δ_{11} plus Δ_{12*} do not add up to total growth. To get total growth one must add an "interaction term."

Experienced growth is unlikely to be the simple sum of the contributions of separate factors. One could take the position that the degree of interaction among the factors is small, and that the separable contributions of the different factors are like the first terms of a Taylor's expansion. This is an arguable position but rests on an assumption about the nature of the production function and about technical change. The approximation might be good, and it might be poor. If the time period in question is considerable, Taylor series arguments are questionable.

The recent literature appears to get around these problems by posing the question in terms of the contribution of average yearly growth of capital, or labor, or technical advance, to average yearly growth of output. However, the contribution of technical change to growth, or of expansion of any particular factor, during any year is not independent of what happened to the other factors prior to that year. Consider a year toward the end of the accounting period. The

contribution of, say, capital during that year is as large as it is because labor and technology had advanced during the prior periods as much as they did. The "average period" metaphor thus may be extremely misleading. And it will be if "interaction" is important.

In principle the questions are answerable. But the answers will require investigations that go beyond growth accounting exercises.

III. What is the Purpose of Growth Accounting?

I take it that the principal purpose of growth accounting has been, and still is, the estimation of various magnitudes which our growth theory suggests are important, testing various propositions or alternative specifications of that theory, and illuminating better the phenomena that a growth theory must explain. Growth accounting is a complement to more traditional (regression) techniques of estimating and testing growth theory. In fact growth accounting has greatly and fruitfully influenced the evolution of our understanding of economic growth. But some of the recent studies seem to imply that somehow the growth accounts really explain growth. I do not see how they can. A growth accounting is not a tested theory of growth.

What do I mean by a tested theory? If we had one, what would it do for us? In the first place we would expect the theory to give an explanation, an account (if not an accounting) of past growth. This immediately poses the question—an account of what phenomena? What needs to be explained? Certainly the aggregative time-series data (at an economy or sector level) on output, input, and prices. Nations and industries have grown at varying rates over time. Different countries and industries have grown at different rates. An explanation is called for.

What do we mean by an explanation? Obviously an explanation implies ability of the theory to replicate reasonably closely the phenomena to be explained, given estimable parameter values. However this is not all we require of a theory. There generally are many alternative explanations that fit past finite data reasonably well. Economists want a theory which will predict as well as explain. And we set premium on the ability of a theory to guide policy. Thus I assume that an acceptable "explanation" does not leave unanswered questions like the two probed in the preceding sections because they imply different things about the future. Prescription as well as prediction hinges on which of these specifications is correct.

In economics, as in all sciences, acceptable explanations must take on certain kinds of analytic forms. Most of growth accounting is motivated by an underlying neoclasical theory of economic growth. Solow (1970) has recently provided a survey of the grab bag of models generally included under this label. Without prejudice regarding the basic question of whether this is the most useful kind of a growth theory, it is important to note that neoclassical theory itself does not answer the above questions. Within the spirit of the theory, there are alternative specifications that answer the questions above in different ways. The questions thus are empirical ones, within the theory.

However, the spirit of growth accounting loads the die. Growth accounting makes "most sense" in a model that assumes that considerable sustained output growth is possible from factor growth, and that factor complementarity is not particularly important. The implicit theory is that output is a function of technology, effective capital, and effective labor. The elasticities of substitution among capital of different qualities, and among labor of different qualities, are assumed to be infinite or at least very high.[6] The elasticity of substitution between effective capital and effective labor implicitly is assumed to be relatively high, and certainly not close to zero. There is considerable ambiguity regarding the connection between technical advance and capital and labor quality. Basically, however, growth accounting as practiced makes sense only if it is assumed that the generation and incorporation of new technology requires only modest amounts of new capital and is not particu-

[6] This is so for models that aggregate capital and which "quality adjust" labor. Not all of the neoclassical models are of this kind.

larly associated with labor of a particular kind or quality.[7]

Under these specifications growth of output per worker can continue so long as the capital-labor ratio grows. While for sustained growth the critical value of the elasticity of substitution is unity, for periods of as long as a quarter century little deceleration of growth of output per worker would be experienced at the rates of factor growth we have experienced, even for an elasticity of substitution as low as one half. Similarly the growth accounting interaction term would not be particularly important over such a time period. Postwar growth of output could be explained quite well as the sum of the separate contributions of improved technology, and increases in effective capital and labor.

However, consider the following alternative model which is at almost an opposite extreme regarding the two questions. Solow, Tobin, von Weizsacker, and Yaari have proposed a model in which any productive increase in the capital intensity of production requires new technology. If technological advance stopped today, output per worker would grow for a while as firms that had been working with older capital shifted over to new and more productive machines. But once they had done this, growth of output per worker would cease. In this model new technology is needed in order to permit productive increase in capital intensity but new technology cannot get into practice without new capital. Let me add to this structure the following. The production and installation of new technology requires educated workers; further, in the absence of technological advance educated workers would be doing nothing different than uneducated workers and would not be more productive.

In this model it is natural to think of technological advance as the binding constraint on the system; certainly sustained growth would be impossible without technical advance. Traditional growth accounting would be nonsense, because of the strong complementarity among technological change, capital growth, and education. To estimate the contribution of technical change by subtracting an estimated contribution of increases in capital and education clearly would be absurd.

The differences between the models involve not only interpretation of past experience, but prescription of how to improve future performance. In the first model it is natural to think of a number of different, and roughly separable, factors that might increase the growth rate. Choice among say more research and development, education, and more physical investment can be made on the basis of rate of return, or cost benefit calculations. While in the long run the complementarity among the factors means that the rate of return on one is not independent of the level of the others, for shorter run calculations this can be ignored. The second model forces policy thinking in terms of complementary packages. Thus a policy in support of *R&D* is thought of as needing support by a policy of training scientists, and as being made effective through policies to facilitate physical investment. An interesting example of a policy which, to be successful, required a rather complex package is the so-called green revolution, as described by Yujiro Hayami and Vernon Ruttan.

To repeat the argument of the earlier sections, we cannot confidently distinguish between these two opposite extremes on the basis of growth accounting exercises and time-series data alone. If we limit ourself to neoclassical formulations, the right model probably is somewhere in between the two cases discussed above, but for many sectors may be closer to the second model than the first. We economists tend to be far too facile with our chalk (or equations) in drawing isoquants into regions of factor proportions that never have been experienced. I would bet that in the absence of considerable research and development reconnaissance of the terrain, firms venturing into technologies with significantly higher capital-labor ratios than actually have been experienced will tend initially to make mistakes, and will experience a considerable amount of learning

[7] For a similar discussion, see Fellner.

costs before achieving significant gains in output per worker. Either research and development (learning before doing) or learning by doing (certainly also a form of *R&D*) is required to make the isoquant more elastic beyond the experienced range. Similarly I believe that economists have been much too mechanical in their treatment of the returns to education. It seems a safe bet that a large share of the returns to higher education are tied up with the processes of technical advance and the chain of economic adjustments thereby set in train.[8]

These are conjectures. We must learn if they are correct, and it is possible to test them. However it strikes me that "growth accounting" is a low marginal returns strategy.

[8] Fellner and Moses Abramowitz and Paul David make similar points. And Hayami and Ruttan make them in their concept of "meta" production function, and their stress on the role of education in innovation and diffusion. Finis Welch discusses these points regarding education, and provides some interesting evidence that education pays off when technology is changing.

REFERENCES

M. Abramowitz and P. David, "Towards Historically Relevant Parables of Growth," Memo. no. 172, Center for Research in Economic Growth, Stanford Univ. 1971.

E. Denison, "Some Major Issues in Productivity Analysis: An Examination of Estimates by Jorgenson and Griliches," *Surv. Curr. Bus.*, May 1969, Part II, *49*, 1–27.

P. Diamond, D. McFadden, and M. Rodriquez, "Identification of the Elasticity of Substitution and the Bias of Technical Change," in D. McFadden, ed., *An Econometric Approach to Production Theory*, Amsterdam 1972.

W. Fellner, "Empirical Support for the Theory of Induced Innovations," *Quart. J. Econ.*, Nov. 1971, *85*, 580–604.

Z. Griliches and D. Jorgenson, "The Explanation of Productivity Change," *Rev. Econ. Stud.*, July 1967, *34*, 249–83.

——— and ———, "The Sources of Economic Growth: A Reply to Edward F. Denison," disc. pap. 131, Harvard Institute of Economic Research, Aug. 1970.

Y. Hayami and V. Ruttan, *Agricultural Development: An International Perspective*, Baltimore 1971.

J. W. Kendrick, *Productivity Trends in the United States*, Princeton 1961.

H. Levine, "A Small Problem in the Analysis of Growth," *Rev. Econ. Statist.*, May 1960, *42*, 225–28.

B. Massell, "Another Small Problem in the Analysis of Growth," *Rev. Econ. Statist.*, Aug. 1962, *44*, 330–32.

M. I. Nadiri, "Some Approaches to the Theory and Measurement of Total Factor Productivity: A Survey," *J. Econ. Lit.*, Dec. 1970, *8*, 1137–77.

M. Richter, "Invariance Axioms and Economic Indexes," *Econometrica*, Oct. 1966, *34*, 739–55.

R. Solow, "Technical Change and the Aggregate Production Function," *Rev. Econ. Statist.*, Aug. 1957, *39*, 312–20.

———, *Growth Theory*, New York 1970.

———, **J. Tobin, C. C. von Weizsacker, and M. Yaari**, "Neoclassical Growth with Fixed Proportions," *Rev. Econ. Stud.*, Apr. 1966, *33*, 79–116.

F. Welch "Education in Production," *J. Polit. Econ.*, Jan.-Feb. 1970, *78*, 35–59.

Faculty Salaries, Promotions, and Productivity at a Large University

By DAVID A. KATZ*

Recently, there has been considerable concern regarding the problems associated with evaluating and rewarding university professors. While these problems have always existed, they have been compounded by the recent economic recession and the growth of political activism which transcends the traditional line between student and teacher.

A recent AAUP report in June 1971 for example cited the "... notable increase in faculty salary grievances, particularly in the form of assertions that a particular salary has been determined upon either for reason violative of academic freedom or on the basis of improper or inadequate evaluation" (p. 187). The report went on to recommend the development of clearly specified and understood criteria for evaluating professors which would hopefully reduce the number of grievances.

Unfortunately, little is known about the process of evaluating and rewarding university professors. One possible reason may be the difficulty of obtaining good data on professors. Secondly, the analysis of such data requires a thorough knowledge of the institutional processes of wage determination. Frequently, it is difficult or impossible to obtain this data and knowledge.

Fortunately, it was possible to obtain the necessary data from a large highly ranked public university for an intensive study of the determinants of faculty salaries and promotions. While it would have been desirable to conduct a multi-university study, this was not possible given the limited time and funds. However, because of the techniques employed in this study, it is expected that the methods can be applied at other large colleges and universities.

I. Methodology

This section presents the methodology which was used to analyze the factors important in salary and promotion decision making at the university level. The primary purpose was to develop a more rational means of evaluating and rewarding university professors.

Personal interviews with the chairmen or heads of nine of the eleven departments in this study (one chairman refused to be interviewed and the other was never available) were conducted to gain a better understanding of the salary and promotion processes. Once budget allocations are made to individual departments, the department chairman sometimes in consultation with his faculty executive committee determines the salaries to be received by individual faculty members. Although final decisions regarding salaries are almost always made at department levels after receiving their budget allocations, department promotional decisions are subject to much greater scrutiny by higher administrative echelons, especially the college and all-university committees on promotions.

Only two of the departments had written policy statements regarding salaries and promotions, available to faculty members

* Assistant professor, department of economics, University of Missouri at Rolla. The research was part of a Ph.D. dissertation in economics, University of Illinois, 1971. I would like to thank the following people: My wife, Debbi, for her considerable help with the tedious task of collecting and coding the massive amount of data; my dissertation advisor, Paul T. Hartman of the University of Illinois, M. Eugene Moyer of the department of Health, Education, and Welfare who helped me considerably in the early stages of the study; John Terwilliger and his assistant Walt Tousey, of the University of Illinois office of administrative studies; David DeVries, now of the Johns Hopkins Center for Study of Social Organization of Schools for allowing me to use the journal ratings constructed by him and Fred E. Fiedler; David Brown, Miami University Provost, for providing the undergraduate school rankings; also, an anonymous referee for his valuable suggestions.

upon request. It will probably surprise no one that the salary and promotion processes within departments are usually cloaked in secrecy.

Very friendly and willing to help, the chairmen and heads explained what criteria they used to judge individuals, but seldom could they provide the relative weights assigned to these criteria. Frequently, they either began or ended by quoting the faculty handbook, supposedly a good source of knowledge regarding salary and promotion policies. "Teaching, research, and public service to the university" were the repeated replies of the chairmen and heads.

After pressing the department administrators on their techniques of evaluating teaching, they usually admitted they knew little about the teaching abilities of professors in their departments, although they thought they knew who were the "good" and "bad" teachers. Without any basis upon which to judge teaching performance, other factors more easily measurable must account for raises in rank and pay. An alternative hypothesis later to be rejected is that salary and promotion decisions were completely arbitrary, with personality and politics playing the major roles.

It became clear that research ability, publication record, and national reputation were the most important factors influencing salary and promotion decisions. The department chairmen and heads agreed they could easily evaluate research ability by examining the quality of journals (usually refereed) in which articles appeared and by reading book reviews. Since department chairmen are seldom experts in every field within their disciplines, this approach is probably better than attempting to judge the merits of research and publication themselves. Especially in large departments, the task of reading and judging everything published would be impossible. Therefore, the quality of an article or book can more easily be ascertained by noting whether or not the article was rated good by experts (as in the usually high quality refereed journals) or if the book received good reviews.

Public service (for example, a public lecture or the editing of a journal), administrative and committee work were the other factors to which department administrators said they gave weight. A current file is kept on all professors containing detailed information on publications, teaching and public service activities, administrative assignments and committee work. It is regularly consulted by department administrators when making salary and promotion decisions.

In the next phase of the study, data from university records were collected which would measure the most important salary and promotion factors. These records contain information on department, years between promotions, date of first appointment, date of birth, degrees and years received, where degrees were received, publications, dissertations supervised, student evaluations of teaching ability, and hours spent per week on committees, public service, administration, teaching, and research. Also, sex, administrative title if any, appointment type (9- or 11-month), percent appointment time, and current salary are included.

As publication records were recorded for each professor, an article which appeared in one of the top journals in any particular discipline was also noted. Recently, David L. DeVries and Fred E. Fiedler developed fairly reliable indices of journal quality in several disciplines. Indices existed for nine of eleven departments in this study and the other two were obtained using methods similar to the DeVries and Fiedler study.

All professors at assistant professor and above in eleven departments were selected for study (596 in total—1969–70 academic year). These departments represent a good cross-section of the major disciplines taught at almost all large colleges and universities—economics, electrical engineering, English, French, history, mathematics, physics, political science, psychology, sociology, and zoology.

II. The Salary Model

The dependent variable is

S = 1969–70 academic year salary corrected to a 100 percent departmental basis.

The independent variables are listed below along with some theoretical justification for inclusion when necessary:

B = number of books published either authored or edited, lifetime.

A = number of articles published, lifetime.

E = number of articles published in the very best journals in a particular discipline ("excellent" articles).

T = a dummy indicating whether or not a professor is ranked in the top 50 percent of all instructors by vote of students.

F = number of dissertations supervised since 1964, a measure of research and graduate teaching ability. Lifetime dissertations were not available.

P = number of hours devoted per week to public service activities in 1970.

C = number of hours devoted per week to committee assignments in 1970.

Y = number of years since highest degree, a measure of experience.

D_1 = a dummy indicating whether or not a professor is a social scientist (economics, political science, or sociology).

D_2 = a dummy indicating whether or not a professor is a physical or laboratory scientist (electrical engineering, mathematics, physics, psychology, or zoology).

D_3 = a dummy indicating whether or not a professor is in the humanities (French or history).

D_4 = a dummy indicating whether or not a professor is in the English department. D_1 to D_4 were included to measure possible differences in market supply and demand for the various major disciplines. D_3 and D_4 were not combined because of considerable salary differences found in previous regression runs.

U_1 = a dummy indicating whether or not a professor received his undergraduate degree from schools containing the top 10 percent of all faculty in the country.

U_2 = a dummy indicating whether or not a professor received his undergraduate degree from schools containing the next 70 percent of all faculty.

U_3 = a dummy indicating whether or not a professor received his undergraduate degree from schools containing the bottom 20 percent of all faculty.

U_4 = a dummy indicating whether or not a professor received his undergraduate degree from a foreign school or one that was not ranked.

G = a dummy indicating whether or not a professor received his highest graduate degree from the bottom groups of graduate school departments. U_1 through G were included to measure quality of education.

M = a dummy indicating whether or not a professor had an administrative assignment.

X = a dummy indicating whether or not a professor was a female. This was included to measure the extent of discrimination, if any.

N = a dummy indicating whether or not a professor was on an 11-month appointment.

R = a dummy indicating whether or not a professor possessed the Ph.D.

The results of this regression are reported in Table 1. Most variables are significant at the 1 percent level and almost all the remaining ones at the 5 percent level. While this regression was capable of explaining 68 percent of the total variation in current salary, the standard error was rather large—$2946.

The following major conclusions can be made regarding salary determination. 1) Teaching ability as measured by student evaluations is inconsequential in the reward system, while publications are paramount. 2) Teaching and research ability at the Ph.D. level, as measured by dissertations supervised, is a most important determinant of salaries. 3) Although public service and committee work are rewarded, the returns are not spectacular. 4) The returns to a graduate education are significantly higher if the degree is received from one of the top graduate school departments. 5) Similarly, it pays not to graduate from one of the bottom ranked undergraduate schools. 6) Social science professors are paid significantly more than those in the humanities, and slightly

TABLE 1—DEPENDENT VARIABLE IS 1969–70 ACADEMIC YEAR SALARY

Variable (Name)	B (Standard Error)	F	Mean	Standard Deviation
Books (B)	230 (86)	7.21***	1.05	2.15
Articles (A)	18 (8)	5.37**	15.36	25.20
Excellent Articles (E)	102 (28)	13.43***	2.73	6.07
Dissertations (F)	489 (60)	66.85***	1.33	2.81
Public Service (P)	89 (38)	5.65**	1.76	3.85
Committees (C)	156 (49)	10.02***	2.90	3.09
Experience (Y)	189 (17)	126.92***	13.13	10.32
Teaching Rating (T)[a]	53 (370)	0.01	0.56	0.50
Physical and Laboratory Scientist (D_2)[b]	−791 (435)	3.31*	0.59	0.49
Humanities (D_3)[b]	−1181 (594)	3.95**	0.10	0.30
English Professors (D_4)[b]	−2293 (529)	18.75***	0.17	0.38
Groups 2, 3, 4, 5 of Undergraduate Schools (U_2)[c]	−474 (364)	1.69	0.57	0.49
Group 6 (U_3)[c]	−2025 (794)	6.50**	0.04	0.19
Foreign and Other (U_4)[c]	27 (478)	0.00	0.17	0.38
Groups 2, 3, 4 of Graduate Schools (G)[d]	−875 (306)	8.15***	0.60	0.49
Administrators (M)[e]	2557 (641)	15.89***	0.05	0.23
Female (X)[f]	−2410 (528)	20.80***	0.08	0.27
11-Month Appointment (N)[g]	1184 (505)	5.50**	0.09	0.28
Ph.D. Degree (R)[h]	1919 (607)	10.01***	0.93	0.25

Constant: 11,155
Multiple $R=.82$ ($R^2=.68$)
$F=50.05$*** Standard Error = 2946
Mean of Dependent Variable = 15,679
Standard Deviation of Dependent Variable = 5093

* Significant at 10 percent level.
** Significant at 5 percent level.
*** Significant at 1 percent level.
[a] Reference group is bottom 50 percent of teachers by vote of students.
[b] Reference group is Social Sciences.
[c] Reference group is top group of undergraduate school rankings.
[d] Reference group is top group of graduate school rankings.
[e] Reference group is those without administrative assignment.
[f] Reference group is male sex.
[g] Reference group is 9-month appointment type.
[h] Reference group is professor without Ph.D.

more than those in the physical and laboratory sciences. 7) Finally, the empirical results tended to strongly support the hypothesis of sex discrimination—women were paid $2410 less in 1969 than men. Each of these conclusions will be discussed more fully below.

Of all variables included in this regression, student evaluations of teachers were the least predictive of salary. While the department chairmen and heads interviewed indicated they rewarded teaching ability, student evaluations were probably not a source of knowledge regarding teaching performance. This conclusion is supported by the fact that most of the department chairmen interviewed had great mistrust of student ratings

In contrast, the publication of a book, article, or excellent article during a professor's lifetime was worth an extra $230, $18, and $102, respectively, in 1969. In a previous regression run, reviews were found to be totally worthless. There were not enough observations on other types of publications to test their significance. The publication regression coefficients represent the salary differentials in 1969 as a result of past publication. Consequently, the actual value of an article or book is measured by discounted

lifetime value which is a function of age.

Another related problem in the interpretation of the regression coefficients is the possibility of nonlinearity. For example, two books or articles may not be equal to twice the value of one. Because of this problem, the linear regression was found to poorly estimate two of the three publication variables—books and articles, but not excellent articles. Nonlinearity was not found to be a serious problem in estimates of the other variable coefficients.

Table 2 presents the average remuneration for various levels of publishing after holding constant all other variables. Diminishing marginal returns to publishing are quite evident in this table. For example, a professor who published one book received $451 more in 1969 than those who had not published any books. However, a professor who published more than two books received an additional $370 per book. Similarly, a professor who published nine to twenty articles received an additional $89 per article in comparison to the under-nine article group which received an additional $111 per article.

TABLE 2—AVERAGE RETURNS TO PUBLISHERS

1–2 Books	$451/book
>2 Books	$370/book
<9 Articles	$111/article
9–20 Articles	$ 89/article
21–38 Articles	$ 74/article
>38 Articles	$ 57/article

In addition to publication record, number of dissertations supervised (since 1964 only) was included as a measure of graduate teaching and research ability and national reputation. Since a dissertation is original research, the supervising of a dissertation requires the same ability which research and publication requires. As proof of this, number of dissertations supervised was highly predictive of current salary (coefficient = $489).

Each additional hour per week of public service or committee work was worth an additional $89 and $156, respectively, in 1969. Assuming a 32-week academic year, this represents a payment of only $2.78/hour for public service and $4.88/hour for committee work.

The graduate school rankings were based on a survey by Allan Cartter of department chairmen and senior and junior professors nominated by deans of 106 graduate schools. The quality of graduate faculties and the effectiveness of doctoral programs were rated by these chairmen and professors. This study used the rated effectiveness of the doctoral program which is grouped into four classes: 1) extremely attractive; 2) attractive; 3) acceptable plus; and 4) not attractive. Because no significant salary differences were found among groups 2–4, they were combined and compared to the top group, extremely attractive. A salary differential of $875 was found to exist in favor of the top group.

The undergraduate school rankings were based on a study by David Brown (1965) which rated 1,121 4-year American colleges and universities. Table 3 presents the groupings of these 1,121 schools. For example, the top 28 schools have 10 percent of the total fulltime faculty but represent only 3.2 percent of the total 1,121 schools, while the bottom 428 schools have only 20 percent of the total faculty but represent 38.2 percent of the 1,121 colleges or universities. Size and quality are therefore strongly correlated in Brown's study.

TABLE 3—UNDERGRADUATE SCHOOL GROUPINGS

Group	Number of Schools in Group	Percent of All Faculty in Group	Percent of All Schools in Group
1	28	10 (top 10)	2.4
2	47	10 (next 10)	4.2
3	134	20 (20–40)	12.1
4	201	20 (40–60)	17.9
5	283	20 (60–80)	25.2
6	428	20 (80–100)	38.2

The 596 professors in this study received their undergraduate educations mostly from the top groups of schools: Group 1, 21 percent; Group 2, 12 percent; Group 3, 23 percent; Group 4, 13 percent; Group 5, 8 percent; Group 6, 4 percent; foreign and other,

16 percent, and missing, 3 percent. The only undergraduate variable which was significant in explaining current salary was Group 6. Apparently, an education at the bottom 428 schools in the country was worth $2025 less than Group 1.

In the early 1960's, a time of excess demand for all disciplines, Brown (1967) ranked 23 disciplines in order of their relative market demands. On a scale of most to least in demand the departments in this study were ranked as follows: electrical engineering (1), mathematics (4), physics (5), economics (6), psychology (10), sociology (11), political science (14), zoology (20), English (21), history (22), and French (23).

Department was therefore included to account for possible differences in market supply and demand for the various disciplines. Social science professors were the best paid after holding constant all other important variables. They earned $791, $1181, and $2293 more than physical-laboratory scientists, humanities (French and history), and English professors, respectively. In the Brown study cited above, English, French, and history were the disciplines least in demand.

The mean 1969 salary for male professors in this study was $16,078, while only $11,233 for women (there were 49 women in the study). After holding constant all other variables, women made $2410 less than men. Therefore, $2410 is a measure of the extent of sex discrimination and $2435 ($16,078 − $11,233 − $2410) is a measure of the salary differential due to other factors, for example, productivity and experience.

An index was constructed to compare the publication records of men and women. This same index was used to compare the publication records of the three ranks (see Table 4). In forming the index the significant publication coefficients (books, articles, and excellent articles) were multiplied by the number of books, articles and excellent articles for each professor. For example, a professor with one book, ten articles and two excellent articles has a "Productivity Index" value of $614=(230\text{x}1)+(18\text{x}10)+(102\text{x}2)$.

The mean Productivity Index value for men was 846 while only 261 for women (difference significant at 1 percent level). This was the clearest indication of the lower productivity of women. Although the following differences were not significant, the mean experience for men in the sample was 13.3 years versus 11.7 years for women; 94 percent of the men had Ph.D.'s versus 88 percent of the women. Men and women had about the same teaching effectiveness ratings.

No significant difference appeared to exist between men and women regarding quality of undergraduate school. In contrast, a significantly higher percentage of women earned their highest degrees from the lower quality graduate school departments: Top Group (39 percent of the men, 22 percent of the women); Second Group (25 percent of the men, 33 percent of the women); Third group (17 percent of the men, 16 percent of the women); Fourth Group (14 percent of the men, 25 percent of the women), and missing (5 percent of the men, 4 percent of the women).

While women devoted fewer hours to public service and administration, these differences were not significant: public service hours per week (men—1.80 hours, women—1.20 hours), administrative hours per week (men—3.83 hours, women—3.46 hours). However, women spent significantly fewer hours on committees (men—3.02 hours, women—1.46 hours).

III. Faculty Promotions and Productivity

To test the extent to which clear standards were applied to promotion decisions, those professors who had received promotions from assistant to associate (249 out of 596 professors) and associate to full professor (183 out of 595 professors) were selected for study. Publication data up to the times of promotions and the total number of years between promotions were available for both groups. Unfortunately, public service and committee data and teaching ratings were recent, and therefore did not accurately measure the extent of involvement in service or committee activities or teaching quality at the times of promotion. If one assumes these do not

TABLE 4—PRODUCTIVITY INDEX BREAKDOWN BY RANK[a]

	0.0	1–100	101–300	301–600	601–1000	1001+
Assistant Professors	22.4	27.1	30.3	15.1	3.0	2.0
Associate Professors	6.4	20.5	27.4	22.0	17.2	6.4
Full Professors	1.6	8.2	15.3	17.5	24.6	32.8

[a] Shown in percent.

change considerably over time as was done with respect to the salary study, then recent values for service, committees, and teaching can serve as proxies for past performance.

In Table 4 the Productivity Index values of professors at the times of promotion to associate and full professors are presented and compared to the records of assistant professors. The index was constructed by multiplying the number of books, articles and excellent articles at the times of promotion (except, of course for assistant professors) by their coefficients in Table 1. The modal value was 101–300 for associate professors; the highest percentage of professors (27.4 percent) was promoted to associate professor after attaining an index value of 101–300. The mean index value was 427 for associate professors and 840 for full professors at the times of promotion in comparison to only 227 for assistant professors (differences significant at 1 percent level).

The fact that the mean Productivity Index values were significantly different among the three ranks supports the conclusion that certain minimal standards for promotion have been practiced. While minimal standards seem to exist, the attainment of this minimum publication record did not guarantee promotion as was indicated by the large number of assistant professors with index values in the 101–300 range (30.3 percent). The mean experience for these assistant professors was only 5 years which suggested that many of them did receive promotions in 1970 given the 6-year tenure rule. (Promotion data is for 1969 in this study.) Therefore, except in unusual circumstances such as extremely high productivity, professors have to wait a certain minimum time before being promoted. This conclusion was supported by the fact that only 21 percent of assistant professors were promoted to associate professors in less than 3 years. The modal value was 3 years, the median 4.8 years, and the mean 5.8 years. Similarly for promotion to full professor, only 17 percent of the associate professors were promoted in less than 3 years; the modal value was 4 years, the median 4.1 years and the mean 4.7 years.

Table 5 compares the rate of productivity publications of associate and full professors before and after promotion. Professors receiving promotions within the last 3 years were excluded because of the possible lag between research and publication. Rate of productivity was calculated by dividing a professor's productivity index by years required to achieve this record.

TABLE 5—RATE OF PRODUCTIVITY COMPARISONS

Rate before promotion to associate professor	=98
Rate after promotion to associate professor	=75
Total Cases=184	
Difference significant at 5 percent level	
Rate before promotion to full professor	=79
Rate after promotion to full professor	=99
Total Cases=138	
Difference significant at 1 percent level	

Productivity declined by about one article per year after promotion to associate professor (difference significant at 5 percent level). However, productivity increased by about the same amount after promotion to full professor (difference significant at 1 percent level). In general, a professor takes a short breather once tenure is granted, but then starts working at his original pace upon attaining the highest rank. Therefore, tenure does not appear to greatly affect lifelong

productivity as measured by publication record.

Teaching ratings did not differ significantly among the ranks. Mean teaching ratings for the three ranks were assistant professor: 5.37; associate professor: 5.42; and full professor: 5.41.

In contrast, public service and committee hours differed significantly among the three ranks. The mean hours per week devoted to public service and committees, respectively, were assistant professor (.92, 1.99), associate professor (1.50, 2.88), and full professor (2.45, 3.51). Therefore, either public service and committees are important variables in explaining rank or some of the perquisites of rank for example are opportunities to lecture to the public, edit a journal, or serve on committees.

The number of hours devoted per week (self-reported) to preparation for teaching did not have any significant effect on a professor's teaching rating. Similarly, the number of hours devoted per week to research had no significant effect on publication record. Professors reported they spent from 0–42 hours in preparation for teaching and 0–70 hours in research. The mean number of hours reported in preparation for teaching was 15 hours, and 19 hours for research.

One of the surprising results of this study was the significant decline in teaching ratings (these are published and distributed to all interested persons) from 1969–70 for those professors with ratings for both years (167 professors). The mean rating for this group was 6.2 in 1969, while only 4.9 in 1970. This may have been a result of changes in the percentile norms from 1969–70. Alternatively, professors may react adversely to published criticism by students.

IV. Recommendations

In general, university administrators make decisions regarding salaries and promotions in an intuitive manner with seldom any clear understanding of the weights they are attaching to various criteria. Consequently, faculty members may have only vague ideas as to how their salaries and promotions are determined. A symptom of the chaotic process is the almost total lack of written departmental policy statements on salaries and promotions. The high standard error of the salary model is further proof of the capriciousness of the reward system.

At present the distribution of rewards among professors in the various departments is arbitrary with no one factor necessarily receiving the most weight. One department may give extraordinary rewards for experience, for example, while another for number of articles published or committee work. Such an uneven distribution of rewards appears to be inequitable. One possible solution might be for the university to establish unambiguous guidelines for departments and colleges which would hopefully result in less arbitrary decisions and fewer grievances.

This study has demonstrated the feasibility of quantifying many of the important determinants of faculty salaries and promotions. If university administrators should decide to institute policy guidelines for departments and colleges to follow, it could serve as a pattern for decision making.

However, some problems still remain before university administrators could adopt this study for use in improving the salary and promotion processes. The foremost of these is the need to define the role of teaching in the reward system. Many administrators and professors feel that the teaching of undergraduates is a minor function at a large university which by its very nature is research and prestige oriented. This view is opposed by a considerable portion of students, administrators, and professors who see the primary role of the university in terms of quality teaching. Since the university in this study is state supported, the final decision in this dispute may rest with the legislative and executive branches, unless university administrators are able to resolve the problem.

If university administrators should decide to institute a teaching rating system of professors, they should be aware of the major problems, especially the great mistrust of student evaluations expressed by so many of the department chairmen and heads. To increase the credibility of teacher evaluations, one possibility might be to poll the

opinions of senior and graduate students only, since presumably they have better knowledge of their particular disciplines than underclassmen. While some have suggested that administrators and colleagues of professors visit classes and seminars to rate faculty members, the results may be worse than senior and graduate student input because the level of tension in having administrators and colleagues visit classes may adversely affect teaching performance.

Administrators should give serious thought to resolving the problems associated with rewarding teaching performance. Otherwise, the state might set standards which would be inconsistent with the goals and purposes of a large college or university, however the latter might be perceived.

Another major area of concern is the salary and rank differential of women versus men. Probably the most important factor accounting for this is the policy of hiring a professor's wife to convince him to accept a job at the university. Supporting this assertion is the fact that about one-half of the women in this study have husbands who are professors at the university and of these about two-thirds have lower rank than their husbands. Consequently, a woman may frequently be considered by administrators as if she were employed on a different basis than her husband, thus reducing her rank and salary.

This study has demonstrated that even after taking into consideration their lower productivity, women are still paid less than men. Although noting the preferential treatment given to hiring women whose husbands are hired at the university, this discriminatory policy in rank and pay is not necessarily justifiable. This complex problem must be resolved by university administrators, or again, as with the case of rewarding teaching effectiveness, the state might step in to solve the problem, but this time in the form of federal action—for example, the cut-off of research grants so important to the functioning of the university.

If the problems of rewarding teaching performance and eliminating sex discrimination are overcome, this study could prove valuable to university administrators. Instead of the present arbitrary and chaotic process of rewarding professors, a more equitable system could be instituted.

REFERENCES

D. G. Brown, "A Report to the Office of Manpower Automation and Training, U.S. Department of Labor," Sept. 1965, vol. 2 in *Academic Labor Markets,* Washington 1965.

———, *The Mobile Professors,* Washington 1967, pp. 12–14.

A. M. Cartter, *An Assessment of Quality in Graduate Education,* Washington 1966.

D. L. DeVries, "The Relationship of Departmental and Personal Role Expectations to the Role Behaviors of University Faculty Members," unpublished doctoral dissertation, Univ. Illinois 1970.

American Association of University Professors Bulletin, "The Role of the Faculty in Budgeting and Salary Matters," June 1971, *57*, 187.

Making Inferences from Controlled Income Maintenance Experiments

By CHARLES E. METCALF*

Since 1968 substantial resources have been devoted to the design and execution of controlled income maintenance experiments.[1] In the typical experiment, a sample of low income households is placed on a variety of negative income tax (*NIT*) plans for a three-year period. The primary objective of these experiments has been to measure 1) the effect of *NIT* on work incentives and 2) the cost of adopting such a welfare scheme nationally.

This paper addresses a question whose answer is essential for interpreting the results of such experiments. If an experiment of limited duration is conducted under "ideal" conditions, what can be inferred from the experimental results about behavior in a world where *NIT* is adopted permanently? By ideal conditions, we mean that (a) individuals (or households)[2] in the experiment are conventional utility maximizers, given the structure of the experiment, and (b) the experiment is sufficiently well designed to permit the differential responses of individuals to be identified.[3] The discussion is limited to a partial equilibrium analysis of an individual's supply of labor (demand for leisure).

Section I extends the usual one-period, static analysis of *NIT* to a multiperiod framework where the effects of a temporary and a permanent *NIT* can be compared. It is shown that, compared to a permanently adopted program, a limited-duration experiment will 1) understate the income effect and 2) overstate the gross and compensated price effects of the *NIT*. Section II analyzes how inferences about behavior under a permanent *NIT* may be drawn from a temporary experiment, and describes how results derived from a two-good, two-period model can be extended to a multigood, multiperiod case. Section III briefly interprets the results outside the context of income maintenance experimentation and outlines some mitigating complications which could affect the results presented in this paper.

I

Consider a one-period model in which a well-behaved function $U(C, L)$ represents the maximum utility an individual may obtain from C dollars of goods purchases and L hours of leisure.[4] The individual takes as given a nonwage income of G dollars, an available quantity of time $\overline{L}$, a wage rate W, and all goods prices. He is assumed to maximize $U(C, L)$ subject to the constraint that $[G+W(\overline{L}-L)-C]$ be nonnegative, i.e., he allocates his consumption between goods and leisure by choosing the quantity of labor he will supply.

* Staff member of the Institute for Research on Poverty and associate professor of economics at the University of Wisconsin. The research reported here was supported by funds granted to the Institute by the Office of Economic Opportunity pursuant to the provisions of the Economic Opportunity Act of 1964. I wish to thank Glen Cain, Robert Haveman, Donald Hester, and Walter Nicholson for their comments on various drafts of this paper. Valuable comments were also received from participants in seminars sponsored by the Institute for Research on Poverty and the Social Systems Research Institute. The conclusions are my responsibility.

[1] In addition to the New Jersey-Pennsylvania Graduated Work Incentive Experiment and the Rural Graduated Work Incentive Experiment (Iowa-North Carolina) funded through the Office of Economic Opportunity, similar experiments are operating in Denver, Gary, and Seattle.

[2] The terms individual and household are used interchangeably in this paper. The household is assumed to possess a single decision maker and to have a finite lifetime corresponding to that of the decision maker.

[3] We are abstracting from all problems related to sample design and stochastic estimation. The analysis of this paper is done in a nonstochastic framework. For a discussion of sample design procedures in such experiments, see Guy Orcutt and Alice Orcutt, and John Conlisk and Harold Watts.

[4] By leisure we mean any nonmarket use of time by the household.

The *NIT* ordinarily provides a household with a fixed income guarantee and simultaneously taxes other income sources of the household at a positive rate up to a "break-even" point where the amount of the guarantee is exhausted. In this model, the *NIT* is assumed to increase nonwage income ($dG>0$) and, through an increase in the tax rate, to decrease the (net of tax) wage rate ($dW<0$).[5] The household is assumed to be below the break-even point, such that its net transfer payment is positive (i.e., $dG+(\bar{L}-L)dW>0$). This latter condition guarantees that the income effect of the guarantee dominates the income effect of the reduction in the wage rate. With a binding budget constraint and the absence of corner solutions (assumed throughout the paper), the imposition of *NIT* increases the demand for leisure through the net income effect if leisure is a normal good. Through the remaining substitution effect, a decline in the wage rate further increases the demand for leisure. Thus, if leisure is a normal good, the *NIT* unambiguously increases the demand for leisure, compared to a situation where no such program is in effect.

In order to compare the effects of a temporary and a permanent *NIT*, the above analysis must be extended to a multiperiod framework. For this exposition consider a two-period, intertemporally additive utility function of the form

$$V = U^1(C_1, L_1) + U^2(C_2, L_2), \tag{1}$$

where the first period coincides with the duration of the temporary *NIT* examined below, and the second corresponds to the remaining lifetime of the individual. The form of $U^1(\cdot)$ and $U^2(\cdot)$ may differ; $U^2(\cdot)$ includes whatever subjective discount factor the individual applies to his future consumption stream.[6] Since the two time periods may be of unequal duration, it should be stressed that G_i, C_i, and L_i ($i=1, 2$) are flows, not aggregates for the time periods in question. The individual maximizes (1) subject to the income constraint

$$\begin{aligned}&[G_1 + W_1(\bar{L} - L_1) - C_1] \\ &\qquad + R[G_2 + W_2(\bar{L} - L_2) - C_2] \geqq 0\end{aligned} \tag{2}$$

where R is the rate at which consumption flows can be transformed from the future to the present. The rate R corresponds to $(1+r)^{-1}$, where r is the one period market rate of interest, if and only if the periods are of equal duration. As period two lengthens relative to period one, R rises to compensate for differences in the time scales; it no longer has a trivial interpretation as an interest rate. If the individual's lifetime is N times the length of period one,[7] and if period two is viewed as N−1 discrete periods each equal to period one in length, then R can be calculated by the formula

$$\begin{aligned}R &= \sum_{i=2}^{N} (1 + r)^{1-i} \\ &= \left[\frac{1 - (1 + r)^{1-N}}{r}\right]\end{aligned} \tag{3}$$

The importance of being able to evaluate R is discussed in Section II.

The conditions for maximizing utility over the two periods are:

$$\begin{aligned} U_C^1 - \lambda &= 0 \\ U_C^2 - \lambda R &= 0 \\ U_L^1 - \lambda W_1 &= 0 \\ U_L^2 - \lambda R W_2 &= 0 \\ [G_1 + W_1(\bar{L} - L_1) &- C_1] \\ + R[G_2 + W_2(\bar{L} - L_2) &- C_2] = 0 \end{aligned} \tag{4}$$

and

$$|H^*| > 0 \tag{5}$$

[5] For expositional purposes, the income guarantee is treated as the only source of nonwage income.

[6] The intertemporal additivity assumption is exploited below in the analysis of substitution effects. It is not required for the results related to income effects. It should also be noted that the individual is assumed to have no bequest motive; he derives utility only from his own consumption.

[7] We can interpret the individual's lifetime as being that portion of his life commencing with the initiation of the *NIT*. Given the form of (1), behavior prior to period one affects behavior during the period of analysis *only* through its effect on the budget constraint.

where $|H^*|$ is the determinant of the relevant Hessian matrix.

A temporary *NIT*, effective during period one, can now be compared with a permanent *NIT*, in force throughout the individual's life. The individual is assumed to know the duration of the *NIT* in each instance, as well as all other relevant information about the future.

The effects of a temporary *NIT* can be derived by changing the values of G_1 and W_1. By differentiating (4) we obtain the effect of a temporary *NIT* upon the first period demand for leisure time.

$$\frac{\partial L_1}{\partial G_1} = \frac{-(W_1 \overset{1}{U}_{CC} - \overset{1}{U}_{LC}) \cdot |D_2|}{|H^*|} \tag{6}$$

and

$$-\frac{\partial L_1}{\partial W_1} = -(\bar{L} - L_1)\frac{\partial L_1}{\partial G_1} - \frac{\lambda}{|H^*|} \cdot [(\overset{1}{U}_{CC} \cdot |J_2|) - |D_2|] \tag{7}$$

where $|D_j|$ is the determinant of the unbordered matrix of second derivatives of $U^j(\cdot)$, and $|J_j|$ is the determinant of the bordered Hessian associated with $|D_j|$.[8] What must be ascertained is whether or not the magnitudes of $(\partial L_1/\partial G_1)$ and the income and substitution components of $(\partial L_1/\partial W_1)$ are unbiased estimates of the magnitudes associated with a permanent change in G and W.

We now differentiate (4) with respect to G and W, where $dG_i = dG$ and $dW_i = dW$, for $i = 1, 2$.

$$\frac{\partial L_1}{\partial G} = (1 + R)\frac{\partial L_1}{\partial G_1} \tag{8}$$

$$\begin{aligned} -\frac{\partial L_1}{\partial W} = &-[(\bar{L} - L_1) + R(\bar{L} - L_2)]\frac{\partial L_1}{\partial G_1} \\ &- \frac{\lambda}{|H^*|}[(\overset{1}{U}_{CC}|J_2|) - |D_2|] \\ &- \frac{\lambda R^2}{|H^*|}(W_1 \overset{1}{U}_{CC} - \overset{1}{U}_{LC}) \\ &\cdot (W_2 \overset{2}{U}_{CC} - \overset{2}{U}_{CL}) \end{aligned} \tag{9}$$

Equation (8) shows the familiar result that consumption is less affected by a transitory change in income than by a permanent change. Evaluated locally, the income effect of a permanent *NIT* is $(1+R)$ times as large as the income effect of a temporary *NIT*. Substituting equation (7) into (9) and simplifying yields:

$$\begin{aligned} -\frac{\partial L_1}{\partial W} = &-\frac{\partial L_1}{\partial W_1} - R(\bar{L} - L_2)\frac{\partial L_1}{\partial G_1} \\ &- \frac{\lambda R^2}{|H^*|}(W_1 \overset{1}{U}_{CC} - \overset{1}{U}_{LC}) \\ &\cdot (W_2 \overset{2}{U}_{CC} - \overset{2}{U}_{CL}) \end{aligned} \tag{10}$$

The final two terms of equation (10) show that the difference in price effects has *both* income and substitution components. If leisure is a normal good, the effect of a reduction in the wage rate on the demand for leisure, estimated in a temporary *NIT* experiment, will typically overstate the permanent price effect for two reasons. First, the estimated negative influence of the income-effect component is biased toward zero. The income effect biases associated with changes in G and W are identical if, and only if, leisure consumption is the same in both periods. Second, the final term in (10) measures the difference between the permanent and transitory substitution effects and will typically be negative.[9]

[8] It should be noted that $|H^*| = |J_1| \cdot |D_2| + |J_2| \cdot |D_1| > 0$. The second-order conditions for optimization also imply that $|J_i| > 0$ for all i, and that $|D_j|$ have a nonpositive value for *at most one j*. If $|D_j|$ is positive for *all j*, an increase in the present value of the individual's income stream will increase total expenditures in every time period. If $|D_j|$ is negative for *one j*, an increase in the present value of income will lead to an expenditure *decrease* in all periods except j. (If $|D_j| = 0$ for *one j*, expenditures would increase in period j and remain constant in all other periods.) In this paper it is assumed that $|D_j| > 0$ for all j, *i.e.*, that the aggregate expenditures in each time period, viewed as a composite commodity, be a "normal good." Nothing in this assumption rules out the possibility of inferiority of individual goods in the consumption bundle. This discussion of restrictions placed on the value of $|D_j|$ parallels the treatment of a single period additive utility function in Eugen Slutsky.

[9] Given that λ and $|H^*|$ are positive, the final term

To summarize this section, if leisure is a normal good, temporary *NIT* experiments will yield estimates of the income effect which is *understated* and of both the gross and the compensated price effects which are *overstated*, relative to what would be observed if a permanent *NIT* program were instituted. It should be noted that a directional statement can be made concerning the *bias* in the gross price effect even though the price effect itself is ambiguous in sign.[10]

II

The identification of three sources of bias (one involving G and two involving W) complicates the process of drawing inferences about permanent behavior from a temporary experiment. In order to make correct inferences the *magnitude* of these biases must be determined. Second, in order to measure the biases, the income and substitution effects must be identified. This implies that different (linearly independent) combinations of income guarantee levels must be included in the design of a *NIT* experiment.

A further examination of equations (8)–(10) suggests procedures for measuring the biases in the estimated income and substitution effects.[11] From (8), the value of R must be known in order to infer the permanent income effect from the estimated transitory measure. Equation (3) indicates that R depends both upon the period interest rate facing the individual and upon his time horizon, measured as a multiple of the length of the experiment. Sample calculations of R illustrate the potential severity of the bias. If the household faces a real annual interest rate of 10 percent ($r=.331$ for a 3-year experimental period) and a time horizon of 30 years (N=10), the bias in the income effect alone would be 2.8 times the measured experimental effect. Even if the time horizon is as short as twice the length of the experiment,[12] this bias would be .75 times the measured effect.

A number of techniques for estimating the value of R are available. If the marginal propensity to consume out of permanent income is known, the result that $1/(1+R)$ equals the ratio of the transitory and permanent marginal propensities to consume can be exploited. Alternatively, if the *NIT* experiment were designed to include subsamples which participated in the program for different time spans, R could be measured directly from observed variations in subgroup income responses.[13]

The income effect component of the bias in the gross price effect (equation (10)) is also a multiple of the observed, temporary income effect, but the value of R does not suffice to measure the size of this component. The present discounted quantity of future labor services provided by the individual must be estimated; in addition to specifying an interest rate, we must assert a relationship between his present and future labor force behavior.

The substitution effect bias in equation (10) can be expressed as:[14]

$$(11)\quad \left[\left(-\frac{\partial L_1}{\partial W}\bigg|_{\bar{U}}\right)-\left(-\frac{\partial L_1}{\partial W_1}\bigg|_{\bar{U}}\right)\right]= -\frac{\lambda R^2}{|H^*|}\,(W_1\overset{1}{U}_{CC}-\overset{1}{U}_{LC})\cdot(W_2\overset{2}{U}_{CC}-\overset{2}{U}_{LC})$$

in (10) will be negative so long as $(W_1U^1_{CC}-U^1_{LC})$ and $(W_2U^2_{CC}-U^2_{LC})$ have the same sign. Given the assumption (see fn. 8) that $|D_j|>0$ and the form of equation (6), $\partial L_i/\partial G_1 \gtrless 0$ as $(W_1U^1_{CC}-U^1_{LC})\lesseqgtr 0$. Thus our requirement is that the effect of an income change on current and future leisure consumption be of the same sign.

[10] If leisure is an inferior good, the income effect would still be understated in absolute value. The statement in fn. 9 about the compensated price effect would stand, but the bias in the gross price effect would be ambiguous in sign.

[11] The entire measurement discussion is in terms of the magnitude of locally evaluated partial derivatives. Additional measurement problems associated with the use of local approximations to evaluate the effects of finite parameter changes are not dealt with in this paper.

[12] While many have argued that low income households have very short time horizons, it is not clear that evidence used to support these arguments in fact support the existence of short time horizons as opposed to high discount rates.

[13] This strategy is being followed in the Seattle-Denver income maintenance experiments.

[14] The notation $|_{\bar{U}}$ is adopted to denote a substitution effect, i.e., the response to a price change accompanied by an income compensation just sufficient to hold utility constant.

This bias can be evaluated by observing the substitution effects of a temporary change in the net wage rate on consumption of all goods during the experiment and estimating the marginal share of leisure in future consumption.

A common adding-up condition from neoclassical consumption theory is that the price-weighted sum of substitution effects of a single price change across all goods equals zero. For a change in W_1, the appropriate condition is:

$$(12)\quad \left(-W_1 \frac{\partial L_1}{\partial W_1}\bigg|_{\bar{U}} - \frac{\partial C_1}{\partial W_1}\bigg|_{\bar{U}}\right) + R\left(-W_1 \frac{\partial L_2}{\partial W_1}\bigg|_{\bar{U}} - \frac{\partial C_2}{\partial W_1}\bigg|_{\bar{U}}\right) = 0$$

Note that the restriction is placed on the price-weighted substitution effects summed over *all* time periods; substitution effects over the period of the experiment are *not* required to sum to zero. The deviation of this latter sum, the first term in (12), from zero is directly related to the substitution bias in (11).

After differentiating (4) to obtain compensated price effects, the weighted sum of $-(\partial L_1/\partial W_1)|_{\bar{U}}$ and $-(\partial C_1/\partial W_1)|_{\bar{U}}$ can be written as:

$$(13)\quad \left(-W_1 \frac{\partial L_1}{\partial W_1}\bigg|_{\bar{U}} - \frac{\partial C_1}{\partial W_1}\bigg|_{\bar{U}}\right) = \frac{-\lambda}{|H^*|}\,(W_1 \overset{1}{U}_{CC} - \overset{1}{U}_{CL}\cdot |J_2|)$$

which is positive or negative depending upon whether leisure is a normal or inferior good during the experimental period.

The marginal share of leisure in future consumption is the remaining element needed to quantify the substitution bias. Again differentiating (4), we can write:

$$(14)\quad R\frac{\partial L_2}{\partial G_1} = \frac{-R^2 |D_1| (W_2 \overset{2}{U}_{CC} - \overset{2}{U}_{LC})}{|H^*|}$$

and

$$(15)\quad R\left(\frac{\partial C_2}{\partial G_1} + W_2 \frac{\partial L_2}{\partial G_1}\right) = \frac{|D_1|\cdot|J_2|}{|H^*|}$$

Taking the ratio of (14) to (15) yields:

$$(16)\quad \frac{R\cdot \dfrac{\partial L_1}{\partial G_1}}{R\cdot\left[\dfrac{\partial C_2}{\partial G_1} + W_2\dfrac{\partial L_2}{\partial G_1}\right]} = \frac{-R^2[W_2 \overset{2}{U}_{CC} - \overset{2}{U}_{LC}]}{|J_2|}$$

Note that the substitution bias defined in (11) equals, with a sign change, the *product* of equations (13) and (16). Thus, from equation (17), we see that the bias in the observed substitution effect is the product of two factors: 1) the weighted sum of substitution effects observed during the experimental period and 2) the marginal share of leisure in future consumption. While the R cancels out of the second factor in this two-period case, discounting plays an important role when the future is disaggregated into multiple periods. Then the second factor becomes the ratio of the effect of an income change on the present discounted *quantity* of future leisure to the effect of an income change on the present discounted *value* of future total consumption.[15] Because the numerator and denominator are measured in different units, the ratio differs from a marginal budget share

Evaluation of the substitution bias requires some estimate of future leisure. One approach would be to assume that present and future leisure shares of total consump-

[15] Specifically, the term corresponding to (16) becomes

$$\frac{\sum_i^{2,N} (1+R)^{1-i}\dfrac{\partial L_i}{\partial G_i}}{\sum_i^{2,N} (1+R)^{1-i}\left(\dfrac{\partial C_i}{\partial G_i} + W_i\dfrac{\partial L_i}{\partial G_i}\right)}$$

$$(17)\quad \left[\left(-\frac{\partial L_1}{\partial W}\bigg|_{\bar{U}}\right) - \left(-\frac{\partial L_1}{\partial W_1}\bigg|_{\bar{U}}\right)\right] = -\left(-W\frac{\partial L_1}{\partial W_1}\bigg|_{\bar{U}} - \frac{\partial C_1}{\partial W_1}\bigg|_{\bar{U}}\right)\cdot\left\{\frac{R\dfrac{\partial L_2}{\partial G_1}}{R\left(\dfrac{\partial C_2}{\partial G_1} + W\dfrac{\partial L_2}{\partial G_1}\right)}\right\}$$

tion are identical at the margin. Given the assumptions underlying this model, an error in the estimated value of the ratio in (17) translates into an equal percentage error in the estimated substitution bias.[16]

The results of this paper can easily be extended to a model with m goods. The reported results concerning income effects can be generalized directly to the multigood case. The biases in the substitution effects of a price change on any two goods can be shown to be proportional to the effects of an income change on consumption of the two goods.[17] If present and future consumption shares (defined as in (16)) are the same for the good undergoing the price change, the price-weighted sum of substitution biases corresponds directly to deviation of the observed sum of substitution effects from zero. Similarly, the results can be reproduced with a multiperiod model in which the form of $U^i(\cdot)$ is again permitted to vary across periods. At each point in the above discussion where future behavior plays a role, the multiperiod analog involves a present discounted quantity or value.

III

The preceding analysis shows how inferences about permanent behavioral relationships can be drawn from measurements of behavior observed in a temporary experiment. The converse problem exists in many discussions of the use of fiscal instruments. Policy makers may wish to predict the effect of temporary fiscal measures from observed empirical effects of earlier fiscal measures which were considered to be permanent. Examples of such instruments would include the 1968 Income Tax Surcharge, temporary investment tax credits, and import surcharges. This paper suggests procedures for analyzing the strength of such fiscal measures; but since it is a partial equilibrium analysis, these procedures are not necessarily reliable for making predictions about the general equilibrium or macro-economic magnitude of their effects.

Several assumptions were employed in the above analysis in order to obtain the reported results. First, the discussion of substitution biases appears to hinge on the assumption that preferences are intertemporally additive. Second, durable and storable goods were excluded from consideration. Third, the households were assumed to be price takers and face perfect capital and labor markets. The nature and degree of imperfection in loan and labor markets serving low income households should be investigated before we either interpret the results of current *NIT* experiments or apply to those results the measures of bias described here. Fourth, households were assumed to adjust instantaneously and costlessly to changes in their economic environment. Again, the magnitude of real world adjustment costs and lags must be investigated. Finally, households were assumed to behave as if they knew the duration of the experiment with certainty. Since this assumption may be violated, and since an experiment may create expectations that *NIT* may be permanently adopted, the effects of uncertainty and of different patterns of expectations on behavior must be examined.

[16] Since consumption patterns in the distant future are discounted heavily, the fact that relative leisure consumption increases substantially after an individual reaches retirement age does not have a major impact on the appropriateness of using current leisure behavior as an approximation of discounted future behavior.

[17] The close relationship between substitution and income effects has its origins in the assumption of intertemporal additivity. For a general discussion of additive preferences, see Hendrik Houthakker.

REFERENCES

J. Conlisk and H. Watts, "A Model for Optimizing Experimental Designs for Estimating Response Surfaces," *Amer. Statist. Assn. Proc., Soc. Statist. Sec.*, 1969, pp. 150–56.

H. S. Houthakker, "Additive Preferences," *Econometrica*, Apr. 1960, *28*, 244–56.

G. H. Orcutt and A. G. Orcutt, "Incentive and Disincentive Experimentation for Income Maintenance Policy Purposes," *Amer. Econ. Rev.*, Sept. 1968, *58*, 754–72.

E. E. Slutsky, "On the Theory of the Budget of the Consumer," *Giornale degli Economisti*, vol. 51, 1915, 1–26; reprinted in J. Stigler and K. Boulding, eds., *Readings in Price Theory*, Homewood, 1952, pp. 27–56.

The Production Function and the Theory of Distributive Shares

By Ryuzo Sato and Tetsunori Koizumi*

One of the stylized facts concerning the performance of an aggregate economy over the last generation or so has been the constancy of income shares between wages and profits. For some this is nothing but the confirmation of the fact that the underlying production function of the economy is of the Cobb-Douglas type, and for others this is due to a bias in technological change with variable elasticity of substitution between labor and capital.[1] But, regardless of whether one believes in an aggregate production function upon which most of these studies are based, or differs in opinion on factors responsible for it, there is a wide consensus among economists on the stability of macro-distributive shares.

The myth surrounding the constancy of factor shares breaks down once we turn our attention to micro units of the economy—at a firm, an industry, or even a sector level. If the production function, which of course becomes better delineated the further we disaggregate, could be expressed in terms of two factor inputs, it would be easy to explain any observed behavior of factor shares by appealing to a simple relation derived by J. R. Hicks between the elasticity of substitution and factor shares.[2] But a micro-production function, since a finer classification of factor inputs is feasible and the data are easier to come by, is conceivably more meaningfully characterized by the multiplicity of factor inputs. The development in the multifactor case, however, had to wait until 1968 when Paul Samuelson first introduced a simple, yet suggestive, concept of the elasticity of substitution. By considering the substitutability of a factor against all other inputs taken together, he was able to demonstrate that the relative share of a factor increases or decreases as the supply of that factor increases depending on whether his elasticity of substitution is greater or smaller than unity. Thus, the usage of the elasticity of substitution in this particular sense was transferred very smoothly from the two-factor to the multifactor case.

Samuelson's new concept, bold and ingenious as it is, has solved only one specific problem, namely, that of the effect of an increase in the supply of *one* factor upon the relative share of that *same* factor. There are, of course, other equally interesting questions one can ask in the multifactor case. For example, what will be the effect of an increase in the supply of *one* factor upon the relative share of *another* factor? What about the effect of a change in the *intensity* of the usage of two factors upon the *ratio* of relative shares of these factors? What about the effect of an increase in the supply of *one* factor upon the *proportion* of relative share of this factor in the total share of a *group* of factors to which this factor may or may not belong? What if the *price* rather than the *quantity* of a factor changes? Not only are these problems of theoretical interest, they are relevant in that they also appear in the real life context as we shall show below with specific examples.

The purpose of this paper is to suggest elements of a complete theory of distributive shares by deriving the different concepts of the elasticity of substitution required in different contexts. Section I briefly discusses the model of production and distribution in the multifactor case. Section II analyzes the role of the Samuelson elasticity and the partial elasticities. Section III discusses the direct and shadow elasticities. In addition, other elasticities are introduced to analyze

* Professor of economics, Brown University and assistant professor of economics at Ohio State University, respectively. This work was supported in part by the National Science Foundation Grant GS-3280. A referee's comments are greatly appreciated.

[1] See, for example, Irving Kravis, Sato.

[2] See Hicks, *Theory of Wages*, ch. 6.

the variety of problems one encounters in real life.

I. The Model

Consider a firm producing output (y) with a n-factor constant returns to scale production function

$$(1) \qquad y = f(x_1, \ldots, x_n)$$

where $x_i = i$th factor input. The function f is assumed to be a concave function with continuous partial derivatives up to, at least, the second order. As dual to the production function (1), we have the unit cost function

$$(2) \qquad c = g(p_1, \ldots, p_n)$$

which is a linear homogeneous, concave function in prices (p_i = price of the ith input). Because of this perfect symmetry between production and cost functions, known as the duality theory, plus the equality between total revenue and total cost which is the equilibrium condition of a firm under pure competition, the subsequent argument on the relationship between the elasticity of substitution and the behavior of factor shares can be conducted in terms of both production and cost functions.

Under competitive conditions, the relative share of factor x_i is given by

$$(3) \qquad \alpha_i = \frac{f_i x_i}{f} = \frac{g_i p_i}{g}, \quad (1 \le i \le n)$$

$$f_i = \frac{\partial f}{\partial x_i}, \quad g_i = \frac{\partial g}{\partial p_i}, \quad \sum_{i=1}^{n} \alpha_i = 1$$

Thus, the relative share can be expressed in two alternative ways either as the proportion of output paid to factor x_i relative to the total output or as the proportion of expenditure on x_i relative to the total cost of production. This enables us to analyze not only the "direct" effect on factor shares of a change in the quantity of a factor via the change in marginal product but also the "indirect" effect of a change in the price of a factor via the change in quantity demanded of that factor.[3]

[3] This is due to the fact that the marginal effect on unit cost of a change in factor price p_i is equal to the quantity demanded of factor i, i.e., $g_i = x_i$.

II. Effects on the Size of Relative Factor Shares

It would be useful to start with a concrete example in order to bring home the empirical relevance of the subsequent analysis. Suppose that the firm under consideration employs two classes of laborers—black and white—in addition to other factor inputs, and that it decides to employ one additional black laborer to fill the job vacancy. How will this affect the relative share of the black laborers? What about its effect on the share of white laborers? In this section, we examine the elasticity concepts required to answer these questions.

Samuelson Elasticity

First consider the effect of an increase in the quantity of *one* factor (black laborer) upon the size of the relative share of *that* factor (black laborer). This is the problem investigated by Samuelson and the result can be stated as: *the relative share of one factor increases or decreases as the quantity of that factor increases depending on whether the "Samuelson" elasticity of substitution is greater or smaller than unity, i.e.,*

$$(4a) \quad sgn\left(\frac{\partial \alpha_i}{\partial x_i}\right) \gtreqless 0 \text{ according as } \sigma_i \gtreqless 1,$$

where

$$(5a) \qquad \sigma_i = \frac{-(1-\alpha_i)f_i}{f_{ii}x_i}$$

is the *Samuelson elasticity of substitution* of factor x_i.

By the aforementioned duality theorem, we can formulate the dual problem to (4a) which is concerned about the effect of an increase in the price of x_i on α_i. The result is given by

$$(4b) \quad sgn\left(\frac{\partial \alpha_i}{\partial p_i}\right) \gtreqless 0 \text{ according as } \sigma_i^d \gtreqless 1$$

where σ_i^d is formally defined as

$$(5b) \qquad \sigma_i^d = \frac{-(1-\alpha_i)g_i}{g_{ii}p_i}$$

We may call σ_i^d in (5b) the "*dual*" *Samuelson*

elasticity of substitution of factor x_i, for it is the exact dual of σ_i in (5a) in that it is defined in terms of the unit cost function (2) which is dual to the production function (1). Thus, (4b) says: *the relative share of one factor increases or decreases as the price of that factor increases depending on whether the "dual" Samuelson elasticity of substitution is greater or smaller than unity.*

Partial Elasticity

Next consider the effect of a change in the quantity (or price) of *one* factor (black laborer) upon the relative share of *another* factor (white laborer). Although the problem seems to be a slight variant of the one just discussed, it turns out that the Samuelson elasticity will not be of much help in answering it. Fortunately for us, however, a rather celebrated "partial" elasticity concept comes to the rescue. The results can be stated as: *the relative share of one factor increases or decreases as the quantity (price) of another factor increases depending on whether the partial elasticity of complementarity (substitution) between the two factors in question is greater or smaller than unity, i.e.,*

$$sgn\left(\frac{\partial\alpha_i}{\partial x_j}\right) \gtreqless 0 \quad \text{according as } c_{ij} \gtreqless 1, \qquad (i \neq j) \tag{6a}$$

$$sgn\left(\frac{\partial\alpha_i}{\partial p_j}\right) \gtreqless 0 \quad \text{according as } \sigma_{ij} \gtreqless 1, \qquad (i \neq j) \tag{6b}$$

where

$$c_{ij} = \frac{ff_{ij}}{f_if_j}, \qquad (i \neq j) \tag{7a}$$

is the *partial elasticity of complementarity* and

$$\sigma_{ij} = \frac{gg_{ij}}{g_ig_j}, \qquad (i \neq j) \tag{7b}$$

is the *partial elasticity of substitution* between x_i and x_j.[4]

Noting that both c_{ij} and σ_{ij} can be either positive or negative (which determines whether x_i and x_j are substitutes or complements),[5] the results in (6a) and (6b) indicate that it is the degree of substitutability which is crucial in determining the behavior of factor shares. For example, suppose that the black and white laborers are strong substitutes, i.e., the partial elasticity of substitution between them is greater than unity. An increase in the wage of white laborers induces the substitution of black laborers for white laborers. An increase in demand for the black laborers thus generated means an increased payment to the black laborers in the total factor payments, thus leading to an increase in the relative share of the black laborers. When the black and white laborers are complementary (i.e., $\sigma_{ij}<0$), an increase in the wage of white laborers always decreases the relative share of black laborers.[6]

III. Effects on the Ratio of Relative Factor Shares

In the previous section we analyzed the effect of a change in the quantity or price of a factor on the size of factor shares. But in the presence of multiple factors equally interesting questions can be asked about the effects on the various ratios of factor shares. Thus, supposing that there is an increase in

[4] The definition of the partial elasticity of "complementarity" is found in Sato and Koizumi (1971) in view of the suggestions due to Hicks (1970). It is the dual concept to the well-known definition of the Allen partial elasticity of substitution.

[5] Two factors x_i and x_j are said to be q-complements or q-substitutes according as c_{ij} is positive or negative, and p-substitutes or p-complements according as σ_{ij} is positive or negative. See Hicks (1970) and Sato and Koizumi (1971).

[6] It is to be noted that the results in (4) and (6) are not totally independent from each other, for the Samuelson elasticity and partial elasticity are related in a certain manner due to the linear homogeneity of f and g functions. To be exact, the relations are

$$1/\sigma_i = \frac{1}{1-\alpha_i}\sum_{j\neq i}\alpha_jc_{ij} \quad \text{and} \quad 1/\sigma_i^d = \frac{1}{1-\alpha_i}\sum_{j\neq i}\alpha_j\sigma_{ij}$$

That is, *the inverse of the Samuelson elasticity of substitution (and that of the dual Samuelson elasticity) of factor x_i is a weighted average of $n-1$ partial elasticities of complementarity (substitution) between x_i and $n-1$ other factor inputs.*

the ratio of black over white laborers, how will this affect the ratio of factor shares between these two classes of laborers? Or one may well conceive of a firm which employs three classes of laborers—black, white, and brown. Then, supposing that a firm decides to increase the use of brown laborers, how will this affect the relative income position of one class of laborers against other classes of laborers? Extending the argument in this manner, one may in general ask about the effect on the ratio of factor shares between two groups of factor inputs. The concepts introduced in the previous section will not help to answer these questions. Hence, we have to develop the elasticity concepts necessary to analyze this type of problem.

Direct and Shadow Elasticities

The simplest case that falls under the present category is the effect of a change in the relative intensity (or relative price) of two factors upon the ratio of factor shares of these factors. As was already pointed out, no new elasticity concept is needed in this case if there are only two factor inputs. In the multifactor case, the results turn out to depend on two other rather celebrated elasticity concepts. That is,

$$\text{(8a)} \quad sgn \frac{\partial\left(\dfrac{\alpha_i}{\alpha_j}\right)}{\partial\left(\dfrac{x_i}{x_j}\right)} \gtreqless 0 \quad \text{according as } d_{ij} \gtreqless 1$$

$$\text{(8b)} \quad sgn \frac{\partial\left(\dfrac{\alpha_i}{\alpha_j}\right)}{\partial\left(\dfrac{p_i}{p_j}\right)} \gtreqless 0 \quad \text{according as } s_{ij} \gtreqless 1$$

where

$$\text{(9a)} \quad d_{ij} = \frac{\partial \log (x_i/x_j)}{\partial \log (f_j/f_i)}, \qquad i \neq j$$

is the *direct elasticity of substitution* between x_i and x_j and

$$\text{(9b)} \quad s_{ij} = \frac{\partial \log (p_i/p_j)}{\partial \log (g_j/g_i)}, \qquad i \neq j$$

is the *shadow elasticity of substitution* between x_i and x_j.[7]

Composite Elasticities

We shall next develop the elasticity concept required to analyze the effect of a change in the quantity (or price) of brown laborers on the *relative income position* of brown laborers against black and white laborers. Let subscripts i, j, and k stand for brown, black, and white laborers, respectively. The income position of brown laborers relative to black and white laborers can be measured by $\alpha_i/(\alpha_j+\alpha_k)$ or its inverse. As the quantity (or price) of brown laborers changes, their relative income position changes as

$$\text{(10a)} \quad sgn \frac{\partial\left(\dfrac{\alpha_j+\alpha_k}{\alpha_i}\right)}{\partial x_i} \gtreqless 0 \quad \text{according as } \psi_i \gtreqless 1$$

$$\text{(10b)} \quad sgn \frac{\partial\left(\dfrac{\alpha_j+\alpha_k}{\alpha_i}\right)}{\partial p_i} \gtreqless 0 \quad \text{according as } \phi_i \gtreqless 1$$

where ψ_i and ϕ_i are the desired elasticities we are looking for. Let us define these elasticities as the *composite elasticities of substitution* of factor i. Formally they are defined as

$$\text{(11a)} \quad \psi_i = (1-\alpha_i)\frac{1}{\sigma_i} + \alpha_i c_i$$

$$\text{(11b)} \quad \phi_i = (1-\alpha_i)\frac{1}{\sigma_i^d} + \alpha_i d_i$$

where

$$c_i = \frac{\alpha_j}{\alpha_j+\alpha_k} c_{ji} + \frac{\alpha_k}{\alpha_j+\alpha_k} c_{ki}$$

[7] The direct elasticity concept is a straightforward extension of the original Joan Robinson definition. The shadow elasticity is due to Daniel McFadden.

TABLE 1—SUMMARY OF ELASTICITY CONCEPTS

Effects on	Changes in x_i	p_i	x_j	p_j	x_i/x_j	p_i/p_j
α_i	Samuelson	Dual Samuelson	Partial elasticity of complementarity	Partial elasticity of substitution		
α_i/α_j					Direct	Shadow
$(\alpha_j+\alpha_k)/(\alpha_i)$	Composite	Dual composite				
$\Sigma\alpha_s/\Sigma\alpha_t$	Group elasticity of complementarity	Group elasticity of substitution				

and[8]

$$d_i = \frac{\alpha_j}{\alpha_j + \alpha_k}\sigma_{ji} + \frac{\alpha_k}{\alpha_j + \alpha_k}\sigma_{ki}.$$

Thus, ψ_i for example is a *convex* combination of the inverse of the Samuelson elasticity and the weighted average of partial elasticities of complementarity between x_i and x_j and between x_i and x_k. Returning to the example of brown versus black and white laborers, the results in (10) indicate that the *increase in the supply of brown laborers is more favorable to their relative income position, the greater is the Samuelson elasticity of substitution of brown laborers against all other factor inputs and/or the smaller are the partial elasticities of complementarity between brown and black and between brown and white laborers.*

Group Elasticities

The composite elasticities discussed above are useful in analyzing the relative income position of one factor against a group of factor inputs. We can, of course, generalize this problem to the comparison of income shares of two groups of factor inputs. Thus, for example, one group of factors may be the workers consisting of black, white, and brown laborers and another group the management consisting of presidents, executives, and branch managers. Supposing that a firm decides to adopt new and efficient capital equipment, how will this affect the relative income positions of these two groups of people? Formally, the results can be stated as

$$\text{(12a)}\quad sgn\,\frac{\partial\left(\sum_{s\in I}\alpha_s \Big/ \sum_{t\in J}\alpha_t\right)}{\partial x_k} \gtreqless 0 \quad \text{according as } C_I - C_J \gtreqless 0$$

$$\text{(12b)}\quad sgn\,\frac{\partial\left(\sum_{s\in I}\alpha_s \Big/ \sum_{t\in J}\alpha_t\right)}{\partial p_k} \gtreqless 0 \quad \text{according as } D_I - D_J \gtreqless 0,$$

where I and J are any two nonintersecting subsets of n factors of production and C_I and C_J are the weighted averages of the partial elasticities of complementarity of group I and J, respectively, and similarly for D_I and D_J with the partial elasticities of substitution. Thus, which group is more favorably affected depends on the relative magnitude of these group elasticities. We may define C_I and D_I as the *group elasticities*

[8] The reader may verify that the same elasticity ψ_i also determines the sign of $\partial[(\alpha_i+\alpha_j+\alpha_k)/\alpha_i]/\partial x_i$. In principle, the composite elasticity could be extended to include more factors. Thus, we may talk about the composite elasticity of brown laborers against black, white, and grey laborers, say, and so on. There is, however, an upper bound on the number of factors to be included. To be exact, that number is $n-2$, for if it is $n-1$ then we would be comparing the income share of one factor against those of all other factors (i.e., α_i versus $1-\alpha_i$) and the relevant elasticity for this problem reduces to the Samuelson elasticity.

of complementarity and substitution of factor group I.[9]

The results we obtained above are schematically summarized in Table 1. It is suggested that the reader fill the remaining blanks with appropriate concepts which will be hybrids of the concepts discussed above.

[9] The results in (10) and (12) are easily obtained by logarithmic differentiation

$$\frac{\partial \log\left(\dfrac{\alpha_j + \alpha_k}{\alpha_i}\right)}{\partial \log x_i} = (1 - \alpha_i)\frac{1}{\sigma_i} + \alpha_i\left(\frac{\alpha_j}{\alpha_j + \alpha_k}c_{ji} + \frac{\alpha_k}{\alpha_j + \alpha_k}c_{ki}\right) - 1$$

and

$$\frac{\partial \log\left(\sum_{s\in I}\alpha_s \Big/ \sum_{t\in J}\alpha_t\right)}{\partial \log x_k} = \alpha_k(C_I - C_J)$$

where C_I, for example, looks like

$$C_I = \frac{\alpha_i}{\alpha_i + \alpha_j + \alpha_l}c_{ik} + \frac{\alpha_j}{\alpha_i + \alpha_j + \alpha_l}c_{jk} + \frac{\alpha_l}{\alpha_i + \alpha_j + \alpha_l}c_{lk}$$

for $I = (i, j, l)$

REFERENCES

R. G. D. Allen, *Mathematical Analysis for Economists*, London 1938.

J. R. Hicks, *Theory of Wages*, 2d ed., New York 1963.

———, "Elasticity of Substitution Again: Substitutes and Complements," *Oxford Econ. Pap.*, Oct. 1970, *22*, 289–96.

I. B. Kravis, "Relative Income Shares in Fact and Theory," *Amer. Econ. Rev.*, Dec. 1959, *49*, 917–49.

D. McFadden, "Constant Elasticity of Substitution Production Functions," *Rev. Econ. Stud.*, June 1963, *30*, 73–83.

J. Robinson, *The Economics of Imperfect Competition*, London 1933.

P. A. Samuelson, "Two Generalizations of the Elasticity of Substitution," in J. N. Wolfe, ed., *Value, Capital and Growth: Essays in Honour of Sir John Hicks*, Chicago 1968.

R. Sato, "The Estimation of Biased Technical Progress and the Production Function," *Int. Econ. Rev.*, June 1970, *11*, 179–208.

——— and T. Koizumi, "Substitutability, Complementarity and the Theory of Derived Demand," *Rev. Econ. Stud.*, Jan. 1970, *37*, 107–18.

——— and ———, "On the Elasticities of Substitution and Complementarity," unpublished mimeo., Brown Univ., Nov. 1971.

R. W. Shephard, *Theory of Cost and Production Functions*, Princeton 1970.

The Neoclassical Theory of Technical Progress: Note

By J. BOHMAN AND M. K. BOHMAN*

In his article on technical progress in this *Review*, Winston Chang presents " . . . the asymmetrical results out of Harrod's and Solow's classifications [of bias] in relation to that of Hicks" (p. 913) for a two-sector model. He also shows " . . . that the usual symmetrical results obtained for aggregate neutrality between the Hicksian and Harrodian schema, . . . in general, do not apply to Solow-neutrality" (p. 913). This note demonstrates that the asymmetries to which Chang refers result from the manner in which he defines the Solow measure of bias in the capital-goods sector. If this measure is redefined, the asymmetries vanish.

Chang defines the Solow parameter, u_{Lj}, $(j=1, 2)$, as the proportional rate of reduction in the labor-output ratio in sector j at a constant wage rate (pp. 913, 918): $u_{Lj} \equiv e_{Lj}|_{\hat{w}=0}$, where $e_{Lj} \equiv (\hat{Y}_j - \hat{L}_j)$; Y_j is the output of sector j, L_j is labor employed in sector j, $(\hat{\ })$ over a variable represents the relative change in that variable, and w is the wage rate measured in terms of the consumption good.[1] It follows that the labor-output ratio under consideration in sector 1 is labor per unit of capital good (L_1/Y_1) whereas the corresponding ratio in sector 2 is labor per unit of consumption good (L_2/Y_2). Chang defines the Solow measure of bias in sector j as $\gamma_j \equiv (1/\theta_{Kj})u_{Lj}$, where θ_{Kj} $(0<\theta_{Kj}<1)$ is the share of capital in sector j.

I. Redefining the Solow Measure of Bias

State the labor-output ratio in sector 1 in terms of consumption goods (L_1/pY_1), where p is the price of a capital good measured in consumption goods. It follows that both the labor-output ratios in sectors 1 and 2 and the wage rate are stated in terms of the consumption good. Denote the proportionate rate of reduction in the labor-output ratios by $\overset{v}{e}_{Lj}$. From the definition of e_{Lj}, it is clear that

$$(1)\qquad \overset{v}{e}_{L1} = \hat{Y}_1 - \hat{L}_1 + \hat{p} = e_{L1} + \hat{p}$$
$$\overset{v}{e}_{L2} = e_{L2}$$

Substitution of Chang's equations (8), (23), and (29) into our equation (1) yields:

$$(2)\qquad \overset{v}{e}_{L1} = (\sigma_1\theta_{K1}/\theta_{K2} + 1 - \theta_{K1}/\theta_{K2})\hat{w} + \theta_{K1}(1-\sigma_1)T_{K2} + \theta_{K1}\beta_1$$

where σ_j, T_{Kj}, and β_j denote the elasticity of factor substitution, the Solow rate of progress, and the Hicks measure of bias, respectively, in sector j. Define the Solow measure of bias, $\bar{\gamma}_j$, such that $\bar{\gamma}_j \equiv (1/\theta_{Kj})\overset{v}{e}_{Lj}|_{\hat{w}=0}$. From our equation (2) and Chang's equation (42), it follows that[2]

$$(3)\qquad \bar{\gamma}_j = \beta_j + (1-\sigma_j)T_{K2}$$

which gives the relationship between the Solow and Hicks measures of bias.[3]

II. Implications

From our equation (3), it follows that Hicks neutrality is equivalent to Solow neutrality in sector j if the elasticity of factor substitution in j equals unity.[4] This result is

[2] γ_j denotes Chang's definition of the measure of Solow bias; $\bar{\gamma}_j$ denotes our definition of that measure. Since no modifications with respect to the consumption-goods sector have been made, $\gamma_2=\bar{\gamma}_2$.

[3] The foregoing procedure of redefining the Solow measure of bias parallels Chang's development of the Harrod measure of bias (pp. 916–17). In the latter case, the capital-output ratio in sector 1 is measured in terms of the capital good. His equation (28), $\overset{*}{e}_{K2}=e_{K2}-\hat{p}$, where $e_{K2}=(\hat{Y}_2-\hat{K}_2)$ and K_2 is the capital employed in sector 2, implies that the relevant capital-output ratio in sector 2 $[pK_2/Y_2=K_2/(Y_2/p)]$ is stated in terms of the capital good. Hence, both the capital-output ratios in each sector and the rate of profit are measured in terms of the capital good.

[4] The above statement holds for sector 2 in Chang's analysis but not for sector 1; see his equation (45).

* University of Santa Clara.

[1] Sector 1 $(j=1)$ is the capital-goods sector; sector 2 $(j=2)$ is the consumption-goods sector.

symmetrical to that of Chang's equation (34) which establishes the relationship between the Harrod and Hicks measures of bias; in particular, the Hicks and Harrod neutralities are equivalent in j if the elasticity of factor substitution in j equals unity.[5] If no progress occurs in sector 2, Hicks neutrality and Solow neutrality are equivalent in sector 1. Contrary to Chang's findings, this result is "... symmetrical to Kennedy's result (1962a, 1962b) concerning the equivalence of Hicks and Harrod neutrality in the consumption-goods sector when there is no progress in the other sector" (p. 919).

Substituting Chang's equation (34) into our equation (3) establishes the relationship between the Harrod and Solow measures of bias:

$$(4) \qquad \bar{\gamma}_j = \alpha_j + (1 - \sigma_j)(T_{L1} + T_{K2})$$

where α_j is the Harrod measure of bias and T_{Lj} is the Harrod rate of progress in sector j. The Solow and Harrod neutralities are equivalent if the elasticity of factor substitution equals unity; hence, if $\sigma_j = 1$, all three classifications of neutrality are equivalent and the production function is Cobb-Douglas.

For the case of Solow own neutrality, define $\bar{\gamma}_1^0 \equiv (1/\theta_{K1})e_{L1}^v|_{\hat{w}-\hat{p}=0}$ and $\bar{\gamma}_2^0 \equiv (1/\theta_{K2})e_{L2}^v|_{\hat{w}=0}$. Clearly, $\bar{\gamma}_2 = \bar{\gamma}_2^0$; hence, substituting Chang's equations (8), (21), and (29) into our equation (1) yields:

$$(5) \qquad \bar{\gamma}_j^0 = \beta_j + (1 - \sigma_j)T_{Kj} + (\theta_{K2}/\theta_{K1})(T_{K2} - T_{Kj})$$

In sector 2, Solow own neutrality is equivalent to Hicks neutrality (or own neutrality) if $\sigma_2 = 1$;[6] on the other hand, in sector 1, the two are equivalent if $\sigma_1 = 1$ and $T_{K1} = T_{K2}$.[7]

Substituting Chang's equation (41) into our equation (5) yields:

$$(6) \qquad \bar{\gamma}_j^0 = \alpha_j^0 + (1 - \sigma_j)(T_{Lj} + T_{Kj}) - (\theta_{L1}/\theta_{L2})(T_{Lj} - T_{L1}) + (\theta_{K2}/\theta_{KI})(T_{K2} - T_{Kj})$$

where θ_{Lj} is labor's share in j and α_j^0 is the Harrod own measure of bias. In sector 1, Harrod own neutrality is equivalent to Solow own neutrality if $\sigma_1 = 1$ and $T_{K1} = T_{K2}$. The same can be said for sector 2 if $\sigma_2 = 1$ and $T_{L1} = T_{L2}$. Both neutralities occur if

$$(7) \qquad \sigma_1 = 1 + (\theta_{K2}/\theta_{K1})(T_{K2} - T_{K1}) \cdot (T_{L1} + T_{K1})^{-1}$$

and[8]

$$\sigma_2 = 1 + (\theta_{L1}/\theta_{L2})(T_{L1} - T_{L2}) \cdot (T_{L2} + T_{K2})^{-1}$$

Following Chang, p. 920, define $\rho \equiv T_{L1}/(T_{L1}+T_{K2})$, where $0 \leqq \rho \leqq 1$. Using Chang's equation (34) and our equation (4), eliminate σ_j. The result is[9]

$$(8) \qquad \beta_j = \rho\bar{\gamma}_j + (1 - \rho)\alpha_j$$

Several symmetrical results emerge which are not present in Chang's analysis (see p. 920). In both sectors, β_j is a weighted average of the other two measures of bias. If $\rho = 0$ (no progress in sector 1), the Hicks and Harrod measures of bias in sector 2 are equivalent and independent of the Solow measure; if $\rho = 1$ (no progress in sector 2),

[5] Chang's equation (34) is as follows:

$$\alpha_j = \beta_j - (1 - \sigma_j)T_{L1}$$

where α_j and T_{Lj} are the Harrod measure of bias and the Harrod rate of progress, respectively, in sector j.

[6] Hicks neutral and Hicks own neutral are the same; see Chang, p. 918, fn. 23.

[7] Contrast this with Chang's equation (48), $\gamma_j^0 = \beta_j + (1-\sigma_j)T_{Kj}$, which indicates that Hicks neutrality and Solow own neutrality are equivalent if $\sigma_j = 1$. Note the symmetry between the results of our equation (5) and that of Chang's equation (41). For the latter, Harrod own neutrality is equivalent to Hicks neutrality in sector 1 if $\sigma_1 = 1$ and in sector 2 if $\sigma_2 = 1$ and $T_{L2} = T_{L1}$.

[8] Compare (6) with Chang's equation (49). In his analysis, the two neutralities are equivalent in sector 1 if $\sigma_1 = 1$; they are equivalent in sector 2 if $\sigma_2 = 1$ and $T_{L1} = T_{L2}$. Both occur if $\sigma_1 = 1$ and

$$\sigma_2 = 1 + (\theta_{L1}/T_{K2})(T_{L1} - T_{L2});$$

see Chang, p. 919.

[9] The corresponding equation in Chang's analysis is his equation (50), $\beta_j = \rho\gamma_j + (1-\rho)\alpha_j + \rho(T_{K2} - T_{Kj})$.

the Hicks and Solow measures in sector 1 are equivalent and independent of Harrod's measure. If progress occurs in both sectors ($0<\rho<1$), then for each sector any two neutralities imply the neutrality of the third.

If Solow (Harrod) neutrality prevails in sector 2 (sector 1), the Hicks classification is equivalent to Harrod's (Solow's).[10] If Harrod (Solow) neutrality prevails in sector 2 (sector 1), Hicks classification, except for the neutrality case, implies the corresponding classification of Solow (Harrod) in that sector, but not vice versa.[11] If Hicks neutrality occurs in sector 1, Solow labor-saving (capital-saving) progress implies Harrod capital-saving (labor-saving) improvement, but not vice versa.[12]

Eliminating σ_j from our equation (5) and Chang's equation (41) yields:

$$(9) \quad \beta_j = \theta_{Kj}\overset{0}{\bar{\gamma}_j} + \theta_{Lj}\overset{0}{\alpha_j} + (\theta_{Lj}\theta_{L1}/\theta_{L2})(T_{L1} - T_{Lj}) + (\theta_{Kj}\theta_{K2}/\theta_{K1})(T_{Kj} - T_{K2})$$

which establishes the relationship among the three measures of own bias.[13] Again several symmetrical results emerge which are not present in Chang's analysis. If the Solow (Harrodian) rates of technical progress between sectors are equal, then: (i) the Hicksian bias in sector 1 (sector 2) is a weighted average of the other two own biases; (ii) any two own neutralities in sector 1 (sector 2) implies the own neutrality of the third; (iii) for sector 1 (sector 2), own neutrality in Harrod's or Solow's sense implies the equivalence of the other two; and (iv) for sector 1 (sector 2), Hicks neutrality implies that Harrod's own labor-saving, neutral, or capital-saving improvement is equivalent to Solow's own capital-saving, neutral, or labor-saving improvement, respectively. If the Solow rate of technical progress in sector 2 is greater (less) than that in sector 1, then for sector 1 (i) Hicks and Harrod own neutrality implies Solow own labor-saving (capital-saving); (ii) Harrod and Solow own neutrality implies Hicks capital-saving (labor-saving); and (iii) Hicks and Solow own neutrality implies Harrod own labor-saving (capital-saving). If the Harrod rate of technical progress in sector 2 is greater (less) than that in sector 1, the same can be said for sector 2.

To reconstruct Chang's formulation of aggregate Solow neutrality, consider his equation (52), p. 921, for the aggregate labor-output ratio:

$$(10) \quad L/Y = \theta_1 L_1/(pY_1) + \theta_2 L_2/Y_2$$

where $L=L_1+L_2$, $Y=pY_1+Y_2$ denotes national income in terms of the consumption good, and θ_j is the share of sector j in Y.

If our equation (4) is substituted for Chang's equation (45), his equation (53) becomes

$$(11) \quad e_L = C_1\hat{w} + C_2 + C_3$$

where

$$C_1 \equiv \lambda_{L2}\sigma_2 + \lambda_{L1}[(\sigma_1 - 1)(\theta_{K1}/\theta_{K2}) + 1]$$
$$C_2 \equiv \lambda_{L1}\theta_{K1}\bar{\gamma}_1 + \lambda_{L2}\theta_{K2}\bar{\gamma}_2$$
$$C_3 \equiv \lambda_{L1}\hat{\theta}_1 + \lambda_{L2}\hat{\theta}_2$$
$$e_L = \hat{Y} - \hat{L}$$
$$\lambda_{Lj} = L_j/L$$

Solow neutrality in each sector ($\bar{\gamma}_1 = \bar{\gamma}_2 = 0$) results in aggregate Solow neutrality ($e_L|_{\hat{w}=0} = 0$) if, and only if, $C_3|_{\hat{w}=0} = 0$. Since, p. 922,

$$(12) \quad \hat{\theta}_1 = \theta_2(1 - \sigma_D)\hat{p}$$
$$\hat{\theta}_2 = -\theta_1(1 - \sigma_D)\hat{p}$$
$$\hat{p}|_{\hat{w}=0} = \theta_{K1}(T_{K2} - T_{K1})$$

it follows that

$$(13) \quad C_3|_{\hat{w}=0} = \theta_{K1}(1 - \sigma_D)(T_{K2} - T_{K1}) \cdot (\lambda_{L2}\theta_1 - \lambda_{L1}\theta_2)$$

[10] In this case, $T_{K2}>0$, $0\leqq\rho<1$, and $\beta_2=(1-\rho)\alpha_2$. For the parenthetical case, $T_{L1}>0$, $0<\rho\leqq 1$, and $\beta_1=\rho\bar{\gamma}_1$.

[11] In this case, $\beta_2=\rho\bar{\gamma}_2$ and $0\leqq\rho<1$. For the parenthetical case, $\beta_1=(1-\rho)\alpha_1$ and $0<\rho\leqq 1$.

[12] The corresponding statement for the other sector is: "... if progress is Hicks-neutral in sector 2, Harrod labor-saving (capital-saving) progress necessarily implies Solow capital-saving (labor-saving) improvement, but again not vice versa." See Chang, p. 920.

[13] Compare with Chang's equation (51), p. 921: $\beta_j=\theta_{Kj}\gamma_j^0+\theta_{Lj}\alpha_j^0+\theta_{L1}(T_{L1}-T_{Lj})$.

where σ_D is the elasticity of substitution between demanders. Solow neutrality in each sector ($\bar{\gamma}_1=\bar{\gamma}_2=0$) implies aggregate Solow neutrality ($e_L|_{\hat{w}=0}=0$) if, and only if, one of the following conditions exists:[14] the elasticity of substitution on the part of demanders equals unity ($\sigma_D=1$) or the Solow rates of technical progress between sectors are equal ($T_{K2}=T_{K1}$).[15] This result parallels Ronald W. Jones' (1966) conclusion that Harrod (Hicks) neutrality in each sector implies aggregate neutrality in the Harrodian (Hicksian) sense if either one of the following conditions exists: the elasticity of substitution on the part of demanders equals unity or the Harrodian (Hicksian) rates of technical progress between sectors are equal.[16]

If sectoral factor proportions differ, then equal Solow, Hicksian, and Harrodian rates of technical progress between sectors cannot all occur. Hence, in the case of unequal sectoral factor proportions, all three classifications of neutrality coexist at both the sectoral and aggregate levels if, and only if, (i) the elasticity of factor substitution in each sector is unitary and (ii) the elasticity of substitution between demanders is unitary.

[14] We ignore the trivial case of equal factor proportions where $(\lambda_{L2}\theta_1-\lambda_{L1}\theta_2)=0$.

[15] According to Chang, "... aggregate Solow neutrality follows if, and only if, either (i) the Solow rate of technical advance, T_{Kj}, is the same in each sector, or (ii) σ_D is greater than unity ..." (p. 922).

[16] Although Jones does not state it, the above conditions are also necessary conditions provided that sectoral factor proportions differ.

REFERENCES

W. W. Chang, "The Neoclassical Theory of Technical Progress," *Amer. Econ. Rev.*, Dec. 1970, *60*, 912–23.

R. W. Jones, "The Structure of Simple General Equilibrium Models," *J. Polit. Econ.*, Dec. 1965, *73*, 557–72.

———, "Comments on Technical Progress," *Philippine Econ. J.*, 1966, no. 2, *5*, 313–32.

C. Kennedy, (1962a) "Harrod on 'Neutrality'," *Econ. J.*, Mar. 1962, *72*, 249–50.

———, (1962b) "The Character of Improvements and of Technical Progress," *Econ. J.*, Dec. 1962, *72*, 899–911.

International Trade and the Forward Exchange Market

By WILFRED ETHIER*

The forward exchange market has recently been the object of much theoretical interest, mostly centered upon the roles played by interest arbitragers and "pure" speculators. Within this framework, discussions of commercial activity assume that all orders are covered completely, so that the forward rate becomes the relevant exchange rate. But an understanding of the interrelationships between trade and forward exchange ought to be a principal objective of studying the forward market. For example, one of the main issues in the question of fixed versus flexible exchange rates concerns the effects of exchange uncertainty on trade. And the rôle of the forward market, both absolutely and comparatively under the two systems, is quite central here. The objective of this paper, then, is to examine the effect of forward exchange on international trade and the relationship between the market for goods and the forward exchange markets.

The present paper differs from most of the recent literature in that the behavior of merchants involves simultaneously both commercial covering and speculation linked in an essential way. Thus there is no deliberate attempt to divide fictitiously the agents into pure speculators, traders, and arbitragers (although one of my objectives is to determine under what circumstances such a division is valid). In this respect my work resembles that of Peter Kenen, to which it is in a sense complementary.[1] Whereas Kenen shows how firms naturally perform the roles of interest arbitragers and speculators in the course of financing their credit operations, I shall focus on the decisions the firm must make at the time of ordering, regarding both the volume of goods to be imported and the amount of forward exchange cover to obtain, and shall ignore completely the way in which the firm actually finances these imports when payment later falls due. I am thus not directly concerned with the familiar leads and lags of commercial payments.

I. The Behavior of the Firm

Opponents of flexible exchange rates typically assert that exchange uncertainty would have a harmful inhibiting effect on world trade. Forward markets, they claim, would not neutralize this because (a) only comparatively few currencies have well-developed forward markets across a wide spectrum of maturities, and (b) even a well-developed forward market cannot eliminate all risk since ". . . changes in the exchange value of currencies are liable to affect merchants in two different ways—through bringing about a change in the amount payable or receivable in terms of their own currency and through bringing about changes in the prices of the goods exported or imported in terms of their own currency" (Paul Einzig, pp. 53–54). Some also claim that (c) forward cover often involves substantial extra cost, or a "risk premium." In this paper I shall ignore completely the first problem and assume that the firm always has recourse to a well-developed forward market for the relevant date.

Consider now the situation of a typical importing firm. Since I wish to examine the implications of exchange uncertainty, I shall assume that the firm possesses certain knowledge about all other facets of its operations. The firm must decide now on the volume M of imports and on the amount of forward cover, supposing that the price of the imports is denominated in foreign currency, and that both the goods will arrive and payment

* Assistant professor of economics, University of Pennsylvania. Thanks are due to Arthur Bloomfield and Stephen Ross for helpful discussions. An early version of this paper was presented to the 1971 New Orleans meeting of the Econometric Society.

[1] See, in addition, William Branson, ch. 2. After the present paper was completed I became aware of the manuscript by Peter Clark which examines some closely related questions from a somewhat different point of view.

will fall due in, say, ninety days.[2] If α denotes the proportion of the import bill for which cover is obtained, the cost in terms of domestic currency for the goods ninety days hence will be $MQ[\alpha R_F+(1-\alpha)R]$ where Q denotes the foreign price of imports (known when the contract is made), R_F the present forward exchange rate, and R the value of the spot rate relevant when payment falls due. The cost in domestic currency of value-added by the firm will be denoted by $V(M)$, and I shall suppose that the firm faces increasing marginal costs, that is, both V' and V'' are positive. I shall assume $V(M)$ is known to the firm at the time it orders, and so independent of any exchange risk. If P denotes the average price in domestic currency which the firm receives for its product, its profit will be given by

$$(1)\quad \pi(M,\alpha) = PM - V(M) - MQ[\alpha R_F + (1-\alpha)R]$$

The firm is free to decide on M and α, and it wishes to make π as large as possible.[3]

The firm is subject to exchange uncertainty both through the domestic cost of its imports (if $\alpha\neq 1$), and through the size of P. The actual way in which P depends upon the exchange rate will be determined by many factors: the market structure in which the firm operates; the policies of its competitors; the sensitivity of demand for its product to the exchange rate; the length of time between orders; the extent to which consumers anticipate price changes; etc. At one extreme, if the firm were in a situation where all competitors ordered at the same time for the same horizon and if demand were insensitive to the exchange rate, P would be determined simply by the aggregate level of industry imports, independently of the behavior of the exchange rate after the goods are ordered. On the other hand it is not difficult to show that in a competitive market where price, marginal cost, and average cost are equal, P will move *pari passu* with R when given time to adjust (which could be very little if customers with perfect foresight quickly react). I shall express the dependence of P on the exchange rate by the hypothesis

$$(2)\quad P = P^0 + \gamma(R - \bar{R})Q$$

where P^0 is a constant, $1\geq\gamma\geq 0$, and $\bar{R}$ denotes the expected value of the random variable R. Three points should be made about this formulation. First there is no reason to restrict attention to competitive firms, so that P (and Q for that matter) might reasonably be considered to be influenced by M as well as by the exchange rate. I have not added this complication since it seems unlikely to yield very much, especially as V already depends on M in a nonlinear fashion. Second, one would expect P to be sensitive not only to the spot rate when payment is due, but to the behavior of the exchange rate from the time the order is first placed until the last unit has been sold. Thus an optimal strategy could involve dealing simultaneously in a spectrum of forward maturities. I have instead adopted the form in (2) because of its simplicity and because it also seems closer to how firms actually behave. A third assumption implicit in (2) is, however, less compelling. By taking γ as a constant known to the firm I am in effect abstracting a situation where firms' uncertainty about exchange rates dominates their uncertainty as to how a change in the exchange rate would affect them. This assumption glosses over an essential aspect of the problem and will therefore be dropped in the next section, but I shall make it now because it does seem reasonable for at least some cases, it leads to particularly simple results, it better enables me to link my results with the standard literature, and it will turn out to be a convenient basis on which to dichotomize the conclusions of this paper.

The firm then will choose M and α in such

[2] If a domestic futures market happens to exist for its product, the firm must also decide on its sales of futures. This will be ignored in what follows.

[3] I am following convention in supposing that the firm is interested in profits in terms of domestic currency rather than in terms of some portfolio of currencies, or some noncurrency standard. As pointed out by Kenen, this can be unrealistic. Also I shall follow convention by ignoring past commitments, which can now be taken as given. This problem is, however, discussed by S. C. Tsiang, and see also Kenen, p. 151.

a way as to maximize $EU(P^0M - V(M) - MQ[\alpha R_F + (1-\alpha)R - \gamma(R-\bar{R})])$, where E denotes mathematical expectation and U is a Von Neumann-Morgenstern utility function. I shall suppose that U is increasing and strictly concave, so that the firm is risk averse in the sense that it will always prefer a certain return to a fair gamble yielding the same expected return. Then if F denotes the distribution function of R and if differentiation can be taken under the integral, the first-order conditions for a maximum give:

$$\frac{\partial EU}{\partial M} = [P^0 - V'(M) - Q(\alpha R_F + \gamma\bar{R})]E(U') - (1-\alpha-\gamma)QE(RU') = 0$$

and

$$\frac{\partial EU}{\partial \alpha} = \int_0^\infty QM(R - R_F)U'(\pi)dF = QM[E(RU') - R_F E(U')] = 0$$

Now the second condition implies $E(RU') = R_F E(U')$ which substituted into the first condition yields

$$(3) \quad V'(M) = P^0 - Q[\gamma\bar{R} + (1-\gamma)R_F]$$

The distinctive feature about (3) is that it does not depend upon the utility function U. The amount M^* that the firm will order (that is, the solution of (3)) is entirely independent of the firm's attitude towards risk, which is relevant only to the choice of the amount of forward cover to obtain. (Of course, M^* may possibly depend upon merchants' overall attitudes towards risk in the general equilibrium sense that the aggregate net demand for cover helps to determine R_F.)

It will be convenient to now look at the typical firm's problem from a mean-standard deviation point of view.[4] That is, I suppose now that the firm looks only at the expected value E_π of profits and at σ, their standard deviation. The risk averse firm prefers higher values of E_π and lower values of σ. It follows from (1) that

$$(4) \quad E_\pi = P^0M - V(M) - [(1-\gamma)R_F + \gamma\bar{R}]QM + (1-\alpha-\gamma)(R_F - \bar{R})QM$$

and

$$(5) \quad \sigma = (1-\alpha-\gamma)QM\sigma_R$$

where σ_R denotes the standard deviation of R.[5]

It is apparent from (5) that the firm can set σ equal to zero by covering to the extent $\alpha = 1-\gamma$. Indeed such a policy will always remove all exchange risk entirely for then the effect of an exchange rate variation on the domestic value of the payment due for imports will always be matched by an equal and opposite effect on the proceeds of domestic sales. It is clear from (4) that if the firm sets $\alpha = 1-\gamma$ it will maximize E_π by also setting $M = M^*$. Then setting α below $1-\gamma$ (if $R_F > \bar{R}$) or above $1-\gamma$ (if $R_F < \bar{R}$) will increase both E_π and σ. Indeed they change in a linear fashion, as substitution of (5) into (4) gives

$$(6) \quad E_\pi = P^0M - V(M) - [(1-\gamma)R_F + \gamma\bar{R}]QM + \frac{R_F - \bar{R}}{\sigma_R}\sigma$$

The fact that U does not appear in (3) implies that such a policy is efficient: if a firm has $M \neq M^*$, then it can *assure* itself of a higher profit by setting $M = M^*$ and covering appropriately. The situation is illustrated in Figure 1. Point A corresponds to $M = M^*$ and $\alpha = 1-\gamma$. The curve $\alpha = 0$ gives the combinations of E_π and σ that correspond to various levels of M if $\alpha = 0$; $M = M^*$ at point B. Similar curves could be drawn for other values of α. The straight line through A and B is the envelope of these curves and the graph of (6) and represents all efficient alternatives; $M = M^*$ everywhere on this line and α falls as one goes from A to B, α being negative beyond B. If $R_F = \bar{R}$ the line AB

[4] The more important results of this section follow from (3) and so do not depend on mean-standard deviation analysis. But its use is expositionally convenient and will facilitate comparison with the next section's results, which are not so easily generalized.

[5] Henceforth in this paper I assume $R_F \geq \bar{R}$. The case $R_F \leq \bar{R}$ can be handled symmetrically.

will be horizontal, so that every risk averse trader will assume a riskless position and cover to the extent $\alpha=1-\gamma$; E_π cannot be increased by assuming added risk.

The above "separation theorem" provides a justification for the conventional method of analyzing the behavior of participants in the forward market. The firm can be thought of as decomposed into a "pure trader" who imports M^* and always covers his order (not completely, as is commonly supposed, but to the extent $1-\gamma$; this distinction will turn out to be significant) and a "pure speculator" who accounts for the difference between α and $1-\gamma$. But a slight qualification is necessary. It is apparent from Figure 1 that the firm's behavior could involve $\alpha<0$, that is, forward sales of foreign exchange. Such behavior would be overtly speculative and so could quite well involve greater cost,[6] that is, the relevant forward rate could be lower than R_F, as α is negative. In this case the efficient frontier becomes $ABCD$. Along AB the firm imports M^* whereas beyond C the size of M equals the solution of (3) for that value of R_F relevant when $\alpha<0$. Between B and C the roles of the volume of imports and amount of cover specified in the separation theorem are precisely reversed: $\alpha=0$ always but the level of M depends upon the firm's attitude towards risk, that is, the level of imports rather than the degree of forward cover furnishes the vehicle for speculation.

Let M' denote the solution of $V'(M)=P^0-\bar{R}Q$. I shall call M' the efficient level of imports since, as is clear from (3), $M^*=M'$ if the forward rate accurately reflects expectations about the future spot rate, that is, if $R_F=\bar{R}$. The possible divergence of the forward rate from the expected future value of the spot rate, as well as from interest parity, is a prominent feature of the modern theory of forward exchange as developed by John Spraos, Sho Chieh Tsiang, Egon

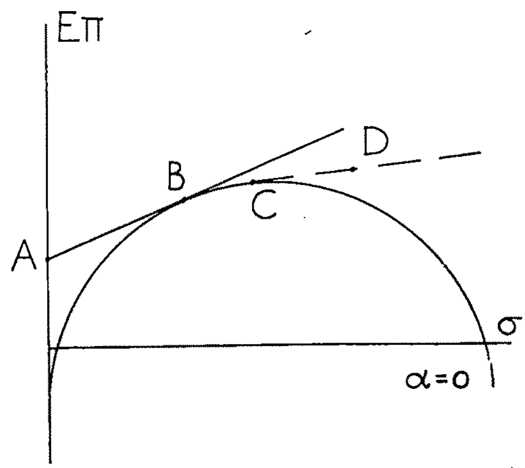


FIGURE 1

Sohmen, and others (see Herbert Grubel for a convenient treatment), and is important for a number of results (see Kenen for example). Merchants frequently ask their bankers whether forward cover currently seems advisable, and in general behave as though they do not take the current forward rate as the best estimate of the future spot rate. Many would argue that market forces would more likely cause R_F and $\bar{R}$ to coincide (with each other as well as with interest parity) under a system of freely floating exchange rates than under one featuring an adjustable peg.[7] Be that as it may, this is quite beyond the scope of the present paper, which will be confined to the implications of the relation between R_F and $\bar{R}$.

Differentiation of (3) leads to

$$(7)\quad \frac{dM^*}{dR_F}=\left(\frac{\partial P^0}{\partial R_F}-(1-\gamma)Q\right)\Big/ V''(M^*)$$

[6] Einzig emphasizes, pp. 104–06, the reluctance of banks to provide forward facilities for overt speculation, and Grubel reports, p. 102, that New York banks require a 10 percent margin deposit. Depending on the size and circumstances of the firm, R_F may or may not be relevant for forward activities that run counter to the firm's trading pattern.

[7] Very briefly, the argument is that if monetary policies should cause domestic and foreign interest rates to differ, then, under a system of floating rates, arbitrage would cause the exchange rate to steadily appreciate or depreciate at a rate to compensate for the difference, thus causing the expected future spot rate, the interest parity level, and the actual forward rate to coincide. Under a pegged system the threat of central bank intervention could make a steady change in the spot rate seem less likely, thus causing the expected future spot rate to differ from the interest parity level, with the actual forward rate falling somewhere in between the two.

Thus, if $R_F > \overline{R}$, M^* will exceed or fall short of the efficient level of imports depending upon whether $\partial P^0/\partial R_F$ is greater than or less than $(1-\gamma)\, Q$.

Sohmen has argued that a premium on forward exchange should not be regarded as a risk premium because "... traders who have assumed commitments to pay in foreign currencies have to pay more for forward than for spot exchange, but other traders who sell their expected foreign exchange receipts forward receive more in exactly the same proportion. For commercial trade as a whole, forward coverage can consequently not be said to be more expensive than spot transactions" (p. 108). If, as Sohmen assumes, all trade is always fully covered, the forward rate is the only relevant exchange rate and, as everyone faces the same forward rate, it can hardly cause distortions in the pattern of trade. But when complete covering is not the universal practice, any departure of R_F from the expected future value of the spot rate necessarily causes inefficiency. This point can be further brought home by also considering a foreign importer of domestic goods who must pay for them on delivery in domestic currency. The analog of (3) for him is

$$W'(X) = P^* - \left[\gamma^* \frac{1}{\overline{R}} + (1-\gamma^*)\frac{1}{R_F}\right] Q^*$$

where X denotes the volume of exports, Q^* their domestic price, W foreign value-added, and P^* and γ^* the foreign counterparts to P^0 and γ. Then the ratio of the domestic price of imports (before value-added so that the goods are comparable internationally) to the domestic price of exports is $[R_F(1-\gamma)+\overline{R}\gamma]\, Q/Q^*$, and the foreign relative price of the same two goods is $Q/Q^*[(1-\gamma^*)/R_F+\gamma^*/\overline{R}]$. If $R_F \neq \overline{R}$ these two relative prices will be equal only in the extreme case where γ and γ^* stand in a certain unique relationship with each other. The usual comparative cost argument then implies that the volume of international trade is not such as to obtain all possible gains. Although the total volume of commercial activity need not be affected by a divergence between the forward rate and the expected future spot rate, in as much as firms selling foreign exchange are influenced in the reverse sense to firms purchasing foreign exchange, as pointed out by Sohmen, the composition of trade will become less efficient. This is because effective relative prices in the two countries will be driven apart when they are sensitive to both R_F and $\overline{R}$.

It is important to realize that this distortion is due entirely to a divergence between the forward rate and the expected future spot rate. Rapid oscillations of these exchange rates, for example, have no such effect provided that the forward rate always reflects expectations about the future spot rate. Also increased uncertainty about the future behavior of R has no effect. An increase in σ_R, for example, will not influence the volume of physical trade but only the amount of forward cover demanded. But will not an increased demand for cover on the part of merchants affect R_F and thus indirectly influence the volume of trade? This is not likely, because domestic and foreign importers and exporters will all demand more cover so that the result will largely consist of an increased volume of forward exchange contracts at the same exchange rate.

It would be of interest to investigate how the demand for imports and for forward cover respond to changes in risk associated with foreign exchange, as represented by σ_R, and to changes in the forward rate. Now it is clear from (3) that the level of imports M^* will be completely unaffected by changes in σ_R, or in any parameter of the distribution function of R other than $\overline{R}$. From (4) and (5) the slope of the line AB in Figure 1 is equal to $(R_F-\overline{R})/\sigma_R$. Thus, since (4) implies E_π is independent of σ_R when $\alpha=1-\gamma$, changes in σ_R result in a pivot of the line AB about the point A, as long as $R_F \neq \overline{R}$. Any risk averse merchant who does not assume the riskless position $\alpha=1-\gamma$ will necessarily be moved to a higher indifference curve by a decline in σ_R, although the level of imports remains unchanged. What about the demand for forward cover? Since M and Q remain unchanged, the demand for forward cover must change in the same proportion as does α.

From (4) E_π will be unaffected by a change in σ_R if α remains constant. It follows therefore that a decline in σ_R will always result in a decline in the demand for forward cover as long as merchants do not display increasing risk aversion to the extreme extent that an increase in wealth is devoted entirely to reducing risk σ and not at all to increasing E_π.[8]

Most writers have assumed that a rise in R_F will cause a fall in both the demand for imports and the demand for cover, and many results in the literature depend on these assumptions.[9] From (7) it follows that a rise in R_F causes a decline in M^* if, and only if, $\partial P^0/\partial R_F<(1-\gamma)Q$, that is, if, and only if, the rise in R_F does not induce a large enough rise in P to increase the profits of a pure merchant assuming the riskless position $\alpha=1-\gamma$. This is true regardless of the firm's actual covering policy. I leave it to the interested reader to investigate the remaining implications of a change in the forward rate.

II. Uncertainty Regarding the Firm's Position

In the previous section I assumed that the typical firm, although uncertain about the future level of the exchange rate, knew what its own profit would be for any value of that rate. This was admittedly unrealistic for many cases, and indeed evades an important issue in the literature. Einzig, for example, states "Even a conservative merchant is not in a position to safeguard himself fully against the indirect effect of the exchange movements, for the simple reason that he is not in a position to foresee the exact extent of that effect. . . . As a result the hedging operation assumes a semi-speculative character" (p. 56). In this section I assume that γ is itself a random variable, distributed independently of R.

[8] Such extreme exceptions will be ruled out if the mean-standard deviation approach is justified by means of a quadratic utility function.

[9] The negative relation between R_F and imports follows from the conventional assumption that all merchants always cover completely (as well as an implicit independence of P from R), but that between R_F and the demand for cover, as Sohmen points out, p. 86, does not.

It follows from (1) that

$$(8)\quad E_\pi = P^0M - V(M) - [(1-\bar{\gamma})R_F + \bar{\gamma}\bar{R}]QM + (1-\alpha-\bar{\gamma})(R_F-\bar{R})QM$$

where $\bar{\gamma}$ denotes the expected value of γ. Also

$$(9)\quad \sigma = QM[(1-\alpha-\bar{\gamma})^2+\sigma_\gamma^2]^{1/2}\sigma_R$$

where σ_γ^2 denotes the variance of γ.

Substitution of (9) into (8) gives

$$(10)\quad E_\pi = P^0M - V(M) - [(1-\bar{\gamma})R_F + \bar{\gamma}\bar{R}]QM + \frac{(1-\alpha-\bar{\gamma})(R_F-\bar{R})}{[(1-\alpha-\bar{\gamma})^2+\sigma_\gamma^2]^{1/2}\sigma_R}\,\sigma$$

In order to delineate the efficient frontier of choices available to the firm, maximize E_π subject to the constraint that σ be equal to some arbitrary value. After some manipulation the first-order conditions for this problem reduce to

$$(11)\quad V'(M) = P^0 - [(1-\bar{\gamma})R_F + \bar{\gamma}\bar{R}]Q - \frac{(R_F-\bar{R})Q\sigma_\gamma^2}{(1-\alpha-\bar{\gamma})}$$

Now the crucial feature of (11), as contrasted with (3), is that it involves the proportion of cover α as well as the level of imports. The separation theorem of Section I is no longer true; the firm's attitude towards risk will help to determine both M and α. Thus when there is uncertainty as to how the individual firm will be affected by the exchange rate it is no longer possible to fictitiously decompose the firm into a pure trader and a pure speculator.

Equations (8), (9), and (11) give three equations in the four unknowns M, α, σ, and E_π. Their solution in terms of E_π and σ is the efficient frontier. Solving these equations ultimately leads to the derivatives shown in equations (12), (13), and (14). Also differentiation of (12) leads to equation (15) which shows that the efficient frontier is strictly concave.

$$\frac{dE_\pi}{d\sigma} = \frac{(R_F - \overline{R})}{\sigma_R} \, \frac{\left[(1 - \alpha - \bar{\gamma})^2 + \sigma_\gamma^2\right]^{1/2}}{(1 - \alpha - \bar{\gamma})} > 0 \tag{12}$$

$$\frac{dM}{d\sigma} = \frac{(R_F - \overline{R})\sigma_\gamma^2\left[(1 - \alpha - \bar{\gamma})^2 + \sigma_\gamma^2\right]^{1/2}}{(R_F - \overline{R})\sigma_\gamma^2 Q \sigma_R\left[(1 - \alpha - \bar{\gamma})^2 + \sigma_\gamma^2\right] + (1 - \alpha - \bar{\gamma})^3 V'' M} > 0 \tag{13}$$

$$\frac{d\alpha}{d\sigma} = \frac{-(1 - \alpha - \bar{\gamma})^2\left[(1 - \alpha - \bar{\gamma})^2 + \sigma_\gamma^2\right]^{1/2} V''}{(R_F - \overline{R})Q^2\sigma_\gamma^2\sigma_R\left[(1 - \alpha - \bar{\gamma})^2 + \sigma_\gamma^2\right] + (1 - \alpha - \bar{\gamma})^3 MQV''} < 0 \tag{14}$$

$$\frac{d^2E_\pi}{d\sigma^2} = \frac{dE_\pi}{d\sigma} \, \frac{\sigma_\gamma^2}{(1 - \alpha - \bar{\gamma})\left[(1 - \alpha - \bar{\gamma})^2 + \sigma_\gamma^2\right]^{1/2}} \, \frac{d\alpha}{d\sigma} < 0 \tag{15}$$

$$\alpha \frac{dM}{d\sigma} + M \frac{d\alpha}{d\sigma} = \frac{\left((1 - \alpha - \bar{\gamma})^2 + \sigma_\gamma^2\right)^{1/2}\left[\alpha(R_F - \overline{R})\sigma_\gamma^2 Q - M(1 - \alpha - \bar{\gamma})^2 V''\right]}{\sigma_R(R_F - \overline{R})Q^2\sigma_\gamma^2\left[(1 - \alpha - \bar{\gamma})^2 + \sigma_\gamma^2\right] + MV''(1 - \alpha - \bar{\gamma})^3 Q} \tag{16}$$

The efficient frontier is depicted by the solid line through C and D in Figure 2. The dotted line through A and B is a reproduction of Figure 1 and represents the efficient frontier in the deterministic case where γ is a parameter with known value equal to $\bar{\gamma}$.

Equations (13) and (14) reveal that a movement along the efficient frontier from C to D and beyond corresponds to an increasing level of imports M and a decreasing proportion of forward cover α. Thus, there is in a sense a tradeoff between mercantile trade and forward cover, with the more speculative firms both importing more and covering less (proportionally) than the more conservative. Will the more speculative firms demand less forward cover absolutely? From (13) and (14), we may derive equation (16), which has the same sign as the numerator term in brackets. If $R_F > \overline{R}$ this term is necessarily positive when M is near zero (and thus α near $1-\bar{\gamma}$) but is necessarily negative when α becomes small. Thus, moving along the curve in Figure 2 from C to D and beyond would cause the total demand for forward exchange to first rise and then fall; the more speculative firm could have either a larger or smaller demand for forward exchange than the more conservative.

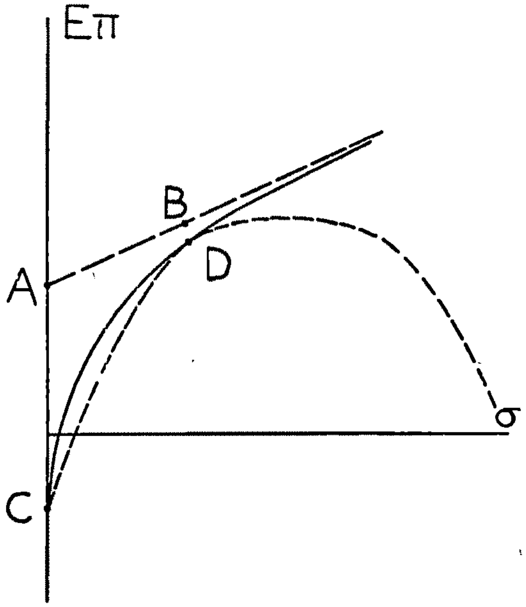


FIGURE 2

Comparison of (11) and (3) reveals that the solution of (11), $M(\alpha)$, is strictly less than M^*, with

$$\lim_{\alpha \to -\infty} M(\alpha) = M^*$$

if (3) reflects the case where γ is assumed to have the value $\bar{\gamma}$ with certainty. Thus, uncertainty about how future profits depend on the future exchange rate must reduce the

level of imports, regardless of the firm's attitude towards risk.[10]

Equation (12) implies that $dE_\pi/d\sigma$ becomes arbitrarily large as α approaches $1-\bar{\gamma}$, and every firm will accept some risk. Indeed the riskless position implies that the firm imports nothing, which entails a certain loss if there are fixed costs. This inability of the merchant to remove all exchange risk is an important distinction between the present treatment and the rest of the literature (including Section I), and is crucial to a consideration of the effects of floating rates on trade.

Comparison of (12) and (8) reveals that the efficient frontier when γ is a random variable is asymptotic, as α becomes small, to the efficient frontier when γ is known with certainty. Thus, if there is serious uncertainty about γ the analysis of Section I gives a good approximation to the behavior of the highly speculative firm, a bad approximation to that of the very conservative firm.

Section I showed that uncertainty about the future exchange rate would reduce the volume of imports by an amount dependent upon the degree to which the forward rate exceeds the expected future spot rate, independently of attitudes toward risk, volatility of the spot rate, etc. Now uncertainty about how the firm's profits depend upon the future spot rate will reduce imports still further, depending upon the willingness of the firm to assume risk. To focus more clearly upon this second factor it is instructive to examine in some detail the situation in the important case where the forward rate accurately reflects expectations about the future value of the spot rate. If $R_F=\bar{R}$ then from (10) it follows that E_π is independent of α, and that E_π will be maximized by ordering M', where $V'(M')=P^0-Q\bar{R}$. Then, from (9), it follows that σ can be minimized for any level of imports by setting $\alpha=1-\bar{\gamma}$. The firm can further reduce risk, at the cost of lower expected profit, by importing less than M'. The resulting efficient frontier is depicted in Figure 3. At point B, $M=M'$ and $\alpha=1-\bar{\gamma}$, while moving from B towards A reduces M, α remaining equal to $1-\bar{\gamma}$.

The situation now is in a sense exactly dual to that of Section I. Every firm will always cover to the extent $\alpha=1-\bar{\gamma}$ whereas the size of its order will reflect its attitude towards risk.

Since M will fall short of M' for conservative firms and will not exceed M' in any case, it is clear that the effect of uncertainty concerning γ will be to reduce the volume of trade. But in contrast to the effect in Section I of a divergence of R_F from $\bar{R}$, this need not be indicative of inefficiency. Uncertainty about γ reflects ignorance of basic economic parameters and should therefore influence the allocation of resources.

Equation (9) implies that $\sigma=QM\sigma_R\sigma_\gamma$ if $R_F=\bar{R}$ and $\alpha=1-\bar{\gamma}$. Thus increasing either σ_γ or σ_R will shift point B horizontally away from the E_π axis, since E_π will not be affected for any given value of M. This will almost certainly reduce the level of imports. (Note also that in this case the effect of such increased uncertainty would be to lower, rather than raise, the forward demand for foreign

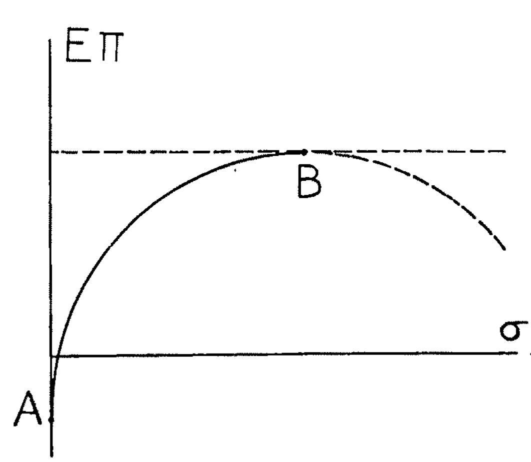


FIGURE 3

[10] Contrast this result with the theory of optimal growth, where uncertainty about the future may either increase or reduce savings, depending upon the response of the degree of risk averseness to changes in wealth. See Leonard Mirman for example. In the case where the expected value of the future spot rate exceeds the current forward rate, movements along the efficient frontier in the direction of higher values of E_π and σ imply that M again increases towards M^* and that α increases beyond $1-\bar{\gamma}$.

exchange.) This analysis could thus be used in an argument against floating rates if one believed that such a system would unjustifiably raise σ_R as a result of destabilizing speculation, etc. Defenders of floating rates, on the other hand, would perhaps assert that such rates would be likely to cause R_F to more closely approximate $\bar{R}$, thus raising efficiency, that σ_R would not likely rise, and that even if it did, the experience of firms under different values of the exchange rate would soon cause a (partially) compensating fall in σ_γ.

This analysis of the firm could be used in a model of the determination of the equilibrium spot and forward rates. If the firm is regarded as typical of the aggregate, the above analysis would give a net demand for forward exchange on the part of trader-speculators as a function of R_F and a net demand for spot exchange determined by past uncovered commitments. The model would be completed by adding a net demand for forward exchange and corresponding supply of spot exchange on the part of interest arbitragers as a function of R_F/R_0 (for given interest rates). There is no point to spelling out such a model in detail here, as it would differ in no essential way from what is already found in the literature (see pp. 97–99 of Sohmen, for example). The equilibrium spot and forward rates simultaneously determined by such a model would turn out to be stable if an increase in the forward rate R_F calls forth an increased net supply of forward foreign exchange. It is worth remarking that the above analysis implies that this condition will almost surely be met near the equilibrium if the equilibrium forward rate closely approximates the expected value of the future spot rate, and if the arbitrage schedule has the conventional shape (see Grubel for example). For an increase of R_F above $\bar{R}$ will almost certainly reduce M and must reduce α below $1-\bar{\gamma}$. But when R_F differs from $\bar{R}$, traders' responses to a change in the forward rate are more ambiguous, as was pointed out in Section I.

III. Concluding Remarks

The foregoing analysis was conducted on the basis of the convention that payment is stipulated in the currency of the exporter. If instead payment were denominated in the importer's currency the basic results would not be affected in any essential way. The exporter would now assume the direct exchange risk whereas the importer would be subject to the indirect exchange risk via the feedback effect on his revenue.

It is possible, however, to also consider the exporter as being subject to a feedback exchange risk. Variations in the exchange rate sometimes affect the cost of production of goods the merchant has contracted to deliver at a stipulated price. For example Einzig claims:

> . . . export trade is not altogether free from the indirect effects of exchange movements, especially in countries whose currencies are susceptible to sufficiently wide fluctuations to react on the internal price level. Thus during the non-stop post-war inflation it was risky for British exporters to sell manufactures for future delivery—even if they covered themselves against the exchange risk by selling forward the currency proceeds—for the simple reason that a devaluation of the pound might cause a sharp rise in the British cost of production in terms of sterling. [p. 54]

But I shall not analyze this situation because it seems to be of much less practical importance than the indirect exchange risk faced by importers and because it also seems unlikely to lead to very different results in any case. Also repercussions of exchange rate variations would cause changes in the production costs of all domestic firms irrespective of whether they were engaged in export trade or not, so it might also be more difficult to construct a model that focused solely on the additional risks faced by a firm engaged in foreign trade.

One might also wish to consider the longer term problem of the effect of exchange risk on the level of investment in firms engaged in foreign trade and on their expected long-run streams of earnings in terms of domestic currency.[11] The present method of analysis can be extended to a formalization of this

[11] See Clark for an interesting investigation along this line.

question without great difficulty, and the basic conclusions do not appear likely to be much affected.

The most important conclusions of this study are as follows. (i) To the extent that firms know how their revenues depend on the future exchange rate, exchange uncertainty influences only the degree of forward cover and not the level of trade. (ii) The pattern of trade will be inefficient to the extent that the forward rate does not accurately reflect merchants' expectations concerning the future behavior of the spot rate. (iii) Uncertainty as to how the firm's revenue depends upon the future exchange rate will cause the level of trade to become sensitive to exchange uncertainty, will reduce the level of trade, and will increase the terms of the tradeoff of expected profit for a reduction in risk. The more speculative the firm, the less significant these factors will be. (iv) If the forward rate accurately reflects merchants' expectations concerning the future spot rate, then firms' attitudes towards risk will be reflected in the level of trade and not at all in the proportion of forward cover. (v) The more closely the forward rate reflects merchants' expectations of the future spot rate the more likely it is that the activities of traders will contribute towards the stability of the spot and forward markets for foreign exchange.

REFERENCES

W. H. Branson, *Financial Capital Flows in the U.S. Balance of Payments,* Amsterdam 1968.

P. B. Clark, "Uncertainty, Exchange Risk, and the Level of International Trade," unpublished manuscript, Duke Univ. 1971.

P. Einzig, *A Dynamic Theory of Forward Exchange,* 2d ed., London 1968.

H. G. Grubel, *Forward Exchange, Speculation, and the International Flow of Capital,* Stanford 1966.

M. C. Kemp, "Towards a Reformulation of the Short-Run Theory of Foreign Exchange," in L. DiMarco, ed., *International Economics and Development: Essays in Honor of Paul Prebisch,* New York 1972.

P. B. Kenen, "Trade, Speculation, and the Forward Exchange Rate," in R. E. Baldwin et al., eds., *Trade, Growth and the Balance of Payments: Essays in Honor of Gottfried Haberler,* Chicago 1965.

L. J. Mirman, "Uncertainty and Optimal Consumption Decisions," *Econometrica,* Jan. 1971, *39,* 179–85.

E. Sohmen, *Flexible Exchange Rates,* rev. ed., Chicago 1969.

J. Spraos, "Speculation, Arbitrage, and Sterling," *Econ. J.,* Mar. 1969, *79,* 1–21.

S. C. Tsiang, "The Theory of Forward Exchange and Effects of Government Intervention on the Forward Exchange Market," *Int. Monet. Fund Staff Pap.,* Apr. 1959, *6,* 75–106.

ANNOUNCEMENT

NOTICE TO ALL GRADUATE DEPARTMENTS

The December 1973 issue of the *Review* will carry the seventieth list of doctoral dissertations in political economy in American universities and colleges. The list will specify doctoral degrees conferred during the academic year terminating June 1973. This announcement is an invitation to send us information for the preparation of the list. This announcement supercedes and replaces a letter which was sent annually from the managing editor's office.

The *Review* will publish in its December 1973 issue the names of those who will have been awarded the doctoral degree since June 1972, the titles of their dissertations, and, if possible, a brief (75-word) summary of the dissertation.

By June 30, please send us this information on 3×5 cards, conforming to the style shown below, one card for each individual. Please indicate by a classification number in the right-hand corner the field in which the thesis should be classified. The classification system is that used by the *Journal of Economic Literature* and printed in every issue.

JEL Classification No. ______

Name: LAST NAME IN CAPS: First Name, Initial

Institution Granting Degree: ______

Degree Conferred (Ph.D. or D.B.A.) ______ Year ______

Dissertation Title: ______

Summary

(75-word maximum, or first 75 words will be printed)

Summary may be completed on back of this card or on new card which should be stapled to this.

When degrees in economics are awarded under different names, such as Business Administration, Public Administration, or Industrial Relations, candidates in these fields whose training has been *primarily in economics* should be included.

All items and information should be sent to the Assistant Editor, *American Economic Review*, Box Q, Brown University, Providence, Rhode Island 02912.

NOTES

Nominations for AEA Officers

The Electoral College on March 2 chose R. A. Gordon as nominee for President-Elect of the American Economic Association in the balloting to be held in the autumn of 1973. Other nominees (chosen by the nomination committee) are: for Vice President (two to be elected), Geoffrey H. Moore, Barbara R. Bergmann, Gary S. Becker, and John G. Gurley; for members of the Executive Committee (two to be elected), Andrew Brimmer, Anne P. Carter, Paul M. Sweezy, and Peter A. Diamond.

Under a change in the bylaws as described in the *Papers and Proceedings* of this *Review*, May 1971, page 472, additional candidates may be nominated by petition, delivered to the Secretary by August 1, including signatures and addresses of not less than 6 percent of the membership of the Association for the office of President-Elect and not less than 4 percent for each of the other offices. For the purpose of circulating petitions, address labels will be made available by the Secretary at cost.

A Group Life Insurance plan for members of the American Economic Association and their dependents is available through the Smith-Sternau Organization of New York. Eligible members may be covered up to $50,000 with $10,000 available under age 50 without medical evidence. A descriptive brochure and enrollment materials will be sent to each member by the Plan Administrator of the Smith-Sternau Organization no later than June 30, 1973.

The fourth Congress of the International Economic Association, to be held in Budapest, Hungary, from August 19–24, 1974, will be open to economists of all countries. The local arrangements committee, expecting a record attendance from both Eastern and Western countries, has reserved 2,400 beds. The program committee is composed of twelve members from as many countries. The general theme of the Congress, *Economic Integration: Worldwide, Regional, Sectoral*, will be treated in three plenary sessions on the first two days, in five sessions each of ten separate working groups, and in two concluding plenary sessions devoted to reports on the proceedings of the working groups. The ten working groups will have the following assignments: Measuring the degree or progress of economic integration; Sectoral Integration: agriculture, transport, energy, selected industries; Industrial Policy: location, technology, multinational firms, competition, and integration of product markets; Migration and integration of labor markets; International capital movements and integration of capital markets; Monetary and Fiscal integration; Socio-Political and institutional aspects of integration; Integration of less developed areas and of areas on different levels of development; Integration by market forces and through planning; World Trade and Interregional Trade: trends and structural changes.

A complete set of *The American Economic Review* from September 1939 to date is available as a donation to an organization, small college, or foreign institution. Please contact Mrs. Paul Webbink, 17 Cohawney Road, Scarsdale, New York 10583.

A Seminar-Workshop on Energy, Growth, and Pollution, sponsored by the departments of geology and economics, Ohio University, will be held July 16–18, 1973. For details, contact R. K. Koshal, 315, Copeland Hall, or Workshop Office, 301, Tupper Hall, Ohio University, Athens, Ohio 45701.

The National Economists Club, an 850-member professional economists organization, held its fifth annual meeting at the International Club of Washington on February 28, 1973. They announce the election of the following officers for the coming year: President, Frank W. Schiff, Committee for Economic Development; Vice Presidents, Nancy H. Teeters, The Brookings Institution and Robert R. Nathan, Robert R. Nathan Associates; Secretary, Grace Taylor, Loomis-Sayles & Company; Treasurer, St. Clair J. Tweedie, American Textile Manufacturers Institute.

New Journal

Research on Consumer Behavior: An interdisciplinary journal will be issued quarterly with the first issue scheduled for July 1974. It will contain two sections: Articles and Communication Notes. Its purpose is to publish empirical research on consumer behavior with special emphasis on articles with interdisciplinary content or interest. For information as to the criteria for articles and communication notes as well as for instructions on manuscript format to be used when submitting articles, write for an *RCB* style sheet to: Ronald E. Frank, editor, *Research on Consumer Behavior*, The Wharton School-W-253, University of Pennsylvania, Philadelphia, Pennsylvania 19174.

A New Series of Economic Research Papers

United Kingdom government economists have commenced publication of their own series of economic research papers under the title *Government Economic Service Occasional Papers*. The aim of the series is to summarize research work which has been undertaken by *UK* government economists in the normal course of their official work but which is also likely to be of interest to economists in general. The *Papers* are published by Her Majesty's Stationery Office. Single copies may be purchased or a standing order for the series as issued placed with the official agent of HMSO in the United States: Pendragon House, Inc., 899 Broadway Avenue, Redwood City, California 94063.

The Volunteer Speakers Service encourages American experts from the professional, academic, and artistic communities to meet with their foreign colleagues or to lecture to select audiences while they are traveling abroad. American scholars and specialists are cordially invited to volunteer as a speaker if they are traveling abroad and have sufficient time at any stop to allow for programming there or in nearby countries. A Volunteer Speaker may be asked to deliver lectures at a university or before other selected audiences, to conduct or participate in a seminar or panel discussion, or perhaps to attend a social event offering an opportunity for informal contact and the exchange of ideas with foreign experts in his field of interest. Interested persons should send their travel plans and professional background details to: Overseas Speakers Division, ICS/S—U.S. Information Agency, Washington, D.C. 20547.

The National Association of Business Economists will hold its next annual meeting at the Plaza Hotel in New York City on September 12–14, 1973. Further details can be obtained from the Arrangements Chairman: John J. Casson, Associate Economist, Brown Brothers Harriman & Co., 59 Wall Street, New York, New York 10005.

Each year since 1954 the Council of the Operations Research Society of America has offered the Lanchester Prize for the best English language published contribution in operations research. The prize for 1972 consists of $2,000 and a commemorative medallion.

The screening of books and papers for the 1972 prize will be carried out by a committee consisting of the three most recent past presidents of the Society. Nominations should be sent to Walter E. Cushen, chairman, 1972 Lanchester Prize Committee, 6910 Maple Avenue, Chevy Chase, Maryland 20015.

The department of economics of the University of Wyoming will sponsor the second annual Bugas Summer School Program in Economics during the 1973 Summer Session. The visiting lecturers for the session include: Ronald G. Cummings, University of Rhode Island, who will offer a course in resource economics and Mohamed A. El-Hodiri, University of Kansas, who will offer a course in mathematical economics. The session will be held June 11–August 3, 1973. The Institute on Health Manpower Planning will be held in August. The exact dates and any information may be obtained by writing the Department of Economics, Box 3985, University Station, Laramie, Wyoming 82071.

The Carolina Population Center of the University of North Carolina is planning a postdoctoral Summer Institute in Population, consisting of workshops for sociologists, economists, psychologists, anthropologists, and political scientists who are interested in research and teaching in the population field, but who need to strengthen and broaden their background in this area. The Institute will be held four weeks, beginning July 30, 1973. Individuals selected to participate will be paid a salary, per diem, and travel. It is anticipated that the Center for Population Research, National Institute of Child Health and Human Development, will support the Summer Institute. Please request details from Dr. Vaida Thompson, director, Summer Institute in Population, Carolina Population Center, University of North Carolina, Chapel Hill, North Carolina 27514. Telephone: 919+933-2055.

The Conference Group on Modern Portugal announces a Workshop on Modern Portugal (1820–1973) to be held October 10–14, 1973 at the New England Center of the University of New Hampshire. Sponsored by a grant from the Council for European Studies, the Workshop will provide an opportunity for social science and history graduate students with an interest in Portugal to meet with American and European scholars currently engaged in studies of social change in Portugal. For details, write Dr. Joyce Riegelhaupt, Institute of Latin American Studies, Columbia University, New York, New York 10027.

A program designed to develop new capabilities in population dynamics has been announced by the Interdisciplinary Communications Program of the Smithsonian Institution. This new effort, called the International Program for Population Analysis and funded by the U.S. Agency for International Development, invites members of the social sciences and humanities—particularly those from the "Third World" of less developed countries—to undertake modest studies of contemporary problems directly related to the development of population policies. Further information may be obtained by writing for the trilingual Publication No. TP-100 (English, French, Spanish) that is available from the Interdisciplinary Communications Program, Smithsonian Institution, 1717 Massachusetts Avenue, N.W., Washington, D.C. 20036.

Deaths

Howard Bourne, associate professor, Wayne State University, Dec. 1, 1972.

Ragnar Frisch, Oslo, Norway, Jan. 31, 1973.

Orris C. Herfindahl, senior research associate, Resources for the Future, Dec. 16, 1972.

N. Arnold Tolles, visiting professor of economics, State University of New York, Geneseo, April 10, 1973.

Leland S. Van Scoyoc, professor emeritus, Bowling Green State University, Jan. 7, 1972.

Retirements

Benjamin Rowe, professor emeritus, social science department New York City Community College of the City University of New York, Sept. 1, 1972.

Clarence W. Tow, senior vice president and director of research, Federal Reserve Bank of Kansas City, Oct. 31, 1972.

Visiting Foreign Scholars

Mrinal Datta-Chaudhuri, Indian Statistical Institute, Delhi: visiting professor of economics, University of Minnesota, Mar. 1973–Mar. 1974.

Assaf Razin, Tel-Aviv University: visiting professor of economics, University of Minnesota, spring quarter, 1973.

Promotions

Charles R. Chittle: associate professor of economics, Bowling Green State University.

Rolf Craft: associate professor of economics and business administration, Luther College.

Raymond J. Doll: vice president and director of research, Federal Reserve Bank of Kansas City, Nov. 1, 1972.

Paul F. Hass: associate professor of economics, Bowling Green State University.

Roger M. Kubarych: special assistant, foreign department, Federal Reserve Bank of New York.

Stanton C. Lindquist: associate professor of economics and business administration, Luther College.

Glenn H. Miller, Jr.: vice president and senior econonist, Federal Reserve Bank of Kansas City, Jan. 2, 1973.

J. David Reed: associate professor of economics, Bowling Green State University.

Sheldon W. Stahl: vice president and senior economist, Federal Reserve Bank of Kansas City, Jan. 2, 1973.

Donald L. Sternitzke: professor of economics, Bowling Green State University.

Frank Zahn: associate professor of economics, Bowling Green State University

Administrative Appointments

Fred Abel, U.S. Department of Agriculture: chief, Economic Analysis Branch, Implementation Research Division, Environmental Protection Agency, Dec. 1971.

Dane J. Cox: treasurer, Colby College, Jan. 30, 1973.

Robert Dockson: chief executive officer, California Federal Savings and Loan Assn., Jan. 17, 1973.

L. Aubrey Drewry, Western Carolina University: dean, College of Business and Applied Science, Marshall University, July 1, 1973.

Paul Gerhardt, National Air Pollution Control Agency: chief, Standards Research Branch, Implementation Research Division, Environmental Protection Agency, Sept. 1971.

Herbert G. Geyer: chairman, department of economics, Hunter College, City University of New York, July 1, 1972.

Molly K. S. C. Lee: chairman, department of social sciences and dean, division of humanities, New York Institute of Technology, Sept. 1, 1972.

F. K. Levy, Rice University: dean, College of Industrial Management, Georgia Institute of Technology, 1972–73.

Geoffrey H. Moore: vice president-research, National Bureau of Economic Research, Feb. 1, 1973.

Robert W. Pullen: administrative vice president, Colby College, Jan. 30, 1973.

H. Leslie Robinson: chairman, department of economics and business administration, Elbert Covell College, University of the Pacific, Dec. 20, 1972.

Marquis Seidel, Back and Sterling, Inc.: chief, Systems Evaluation Branch, Implementation Research Division, Environmental Protection Agency, Apr. 1972.

Gloria M. Shatto, University of Houston: associate dean, College of Industrial Management, Georgia Institute of Technology, 1972–73.

Richard D. Teach: associate dean, College of Industrial Management, Georgia Institute of Technology, 1972–73.

Appointments

John Adams: instructor in economics and business administration, Luther College.

Fred C. Allvine, Northwestern University: associate professor of industrial management, Georgia Institute of Technology, 1972–73.

Scott Atkinson, University of Colorado: economist, Implementation Research Division, Environmental Protection Agency, Aug. 1972.

Uri Ben-Zion, University of Chicago: visiting professor of economics, University of Minnesota, Sept. 1972–June 1973.

Brian Bock: instructor in economics, Luther College.

Patricia Brenner: assistant professor, department of economics, Williams College, Sept. 1972.

John Ciccolo: economist, Money and Finance Division, Federal Reserve Bank of New York.

James F. Courtney: instructor of industrial management, Georgia Institute of Technology, 1972–73.

David A. Couts: coordinator of research, Motor Vehicle Manufacturers Assn., Detroit.

John Danforth, Northwestern University: assistant professor of economics, University of Minnesota, Sept. 1972.

Mark Z. Fabrycy, New York University: professor of economics, Wright State University, Jan. 1973.

Elon Gilbert: visiting scholar, Food Research Institute, Stanford University.

James R. Hibbs, U.S. Department of Commerce: assistant to director, Implementation Research Division, Environmental Protection Agency, Dec. 1972.

W. Whitney Hicks: visiting associate professor, Food Research Institute, Stanford University.

John H. Hoag: assistant professor of economics, Bowling Green University.

Martin E. Horning: instructor of economics, Bowling Green University.

Robert J. Jones, Columbia University: instructor in economics, Skidmore College, Feb. 1973.

Jay H. Levin, Federal Reserve Board: associate professor of economics, Wayne State University, Jan. 1, 1973.

Wuu-Long Lin: research associate, Food Research Institute, Stanford University.

John McCracken: economist, Foreign Research Division, Federal Reserve Bank of New York.

Edward F. McKelvey: assistant professor, department of economics, Williams College, Sept. 1972.

Bridger M. Mitchell, Department of Health, Education, and Welfare: staff member, economics department, The Rand Corporation, Nov. 1972.

Marcelo Peinado, Utah State University: associate professor of agricultural economics, Elbert Covell College, University of the Pacific, Aug. 16, 1972.

Henry M. Peskin: research associate, National Bureau of Economics Research, Dec. 1972–Dec. 1973.

James L. Plummer, U.S. Embassy, Bogota, Colombia: economist, Environmental Protection Agency, Implementation Research Division, July 1972.

Robert Reck, University of Maryland: economist, Implementation Research Division, Environmental Protection Agency, Dec. 1971.

Marshall Rose, Xerox Corporation: economist, Implementation Research Division, Environmental Protection Agency, June 1972.

David K. Sandberg: economist, Balance of Payments Division, Federal Reserve Bank of New York.

Richard A. Sandell: associate professor of international business and economics, Elbert Covell College, University of the Pacific, Aug. 16, 1972.

Peter G. Sassone: assistant professor of economics, Georgia Institute of Technology, 1972–73.

Rudolph K. Schnabel: instructor, department of economics, Iowa State University.

Paul T. Schultz, The Rand Corporation: professor of economics, University of Minnesota, Sept. 1972.

David L. Scott: assistant professor of economics, Florida Southern College, Sept. 1972.

Robert J. Shiller, Massachusetts Institute of Technology: assistant professor of economics, University of Minnesota, Sept. 1972.

William L. Stringer: assistant professor of economics, Bowling Green State University.

Dennis Tihansky, The Rand Corporation: operations research analyst, Implementation Research Division, Environmental Protection Agency, Jan. 1972.

Henry C. Wallich: senior research associate, National Bureau of Economic Research.

William Watson, Washington State University: economist, Implementation Research Divisi mental Protection Agency, June 1972.

Leaves for Special Appointn

Garry M. Crane, National Bureau of Sta ing economist, Department of Supply Canada, Sept. 1972.

Paul B. Downing, University of Californ economist, Implementation Research Di ronmental Protection Agency, July 1972–

B. F. Kiker, University of South Car professor, University of Edinburgh, Scotl 1973.

George von Furstenberg, Indiana Unive professor of economics, University of Au many, fall semester 1972.

Gary M Walton, Indiana University: v sor of economics, University of Hawaii, sp 1973.

Resignations

Irving Auerbach, Federal Reserve Bank Aubrey G. Lanston & Co., Inc.

John Buttrick, University of Minnesot versity, June 1973.

Peter Gregory, University of Minnesot of New Mexico, June 1973.

Ralph Hofmeister, University of Mi World Bank June 1972.

David T. Lapkin, University of No Chapel Hill, Dec. 31, 1972.

Richard G. Marcis, Bowling Green Stat Federal National Mortgage Assn., July 19

Allan P. Rahn, Iowa State University: Georgia.

David E Vanderford, Federal Reserve York: Townsend & Greenspan, Inc.

NOTE TO DEPARTMENTAL SECRETARIES AND EXECUTIVE OFFICERS

When sending information to the *Review* for inclusion in the Notes Section, please use the following

A. Please use the following categories:

1—Deaths
2—Retirements
3—Foreign Scholars (visiting the USA or Canada)
4—Promotions
5—Administrative Appointments
6—New Appointments
7—Leaves for Special Appointments (NOT
8—Resignations
9—Miscellaneous

B. Please give the name of the individual (SMITH, John W.), his present place of employment or er new title (if any), and the date at which the change will occur.

C. Type each item on a separate 3×5 card and please do not send public relations releases.

D. The closing dates for each issue are as follows: *March*, November 1; *June*, February 1; *Septer December*, August 1.

This announcement supersedes and replaces a letter which was sent annually from the managing editc items and information should be sent to the Assistant Editor, *American Economic Review*, Box Q, Brow Providence, Rhode Island 02912.

The American Economic Review

ARTICLES

SEPTEMBER 1973

THE AMERICAN ECONOMIC ASSOCIATION

Founded in 1885

• Published at George Banta Co., Inc., Menasha, Wisconsin.

• THE AMERICAN ECONOMIC REVIEW, including four quarterly numbers, the *Proceedings* of the annual meetings, and *Directory* and Supplements, is published by the American Economic Association and is sent to all members five times a year, in March, May, June, September, and December.

• Membership dues of the Association are $21.00 a year, which includes a year's subscription to both the *American Economic Review* and the *Journal of Economic Literature*. A three year junior membership of $10.50 per year is available to registered students. Subscriptions by nonmembers are $31.50 a year, and only subscriptions to both publications will be accepted. Single copies of the *Review* and *Journal* are $4.50 each. Each order for copies of either publication must also include a $.50 per order service charge. Orders should be sent to the Secretary's office, Nashville, Tennessee.

• Correspondence relating to the *Papers and Proceedings*, the *Directory*, advertising, permission to quote, business matters, subscriptions, membership and changes of address may be sent to the secretary, Rendigs Fels, 1313 21st Avenue, South, Nashville, Tennessee 37212. To be effective, notice of change of address must reach the secretary by the 1st of the month previous to the month of publication. The Association's publications are mailed by second class and are not forwardable by the Post Office.

• Second-class postage paid at Nashville, Tennessee and at additional mailing offices. Printed in U.S.A.

R. G. Gordon

R. A. GORDON

Distinguished Fellow
1972

In honoring Robert Aaron Gordon, we applaud craftsmanship in the service of compassion. From his early books *Business Leadership in the Large Corporation* and *Business Fluctuations* to his recent work on the structure of labor markets, culminating in *The Goal of Full Employment*, he has steadfastly pursued two objectives for economics: that it should fit the facts of modern economic life, and that it should find ways to improve the human condition. In a time when it was not fashionable, he has been the practical man's theorist, the theorist's practical man, and everybody's enemy of stereotypes. His contributions to economics include his institutional work on the control of large corporations, his theoretical and empirical analysis of business cycles, and his close study of manpower problems. For all of this, we thank him.

The 1973 Report of the President's Council of Economic Advisers: The Economic Role of Women

By Barbara R. Bergmann and Irma Adelman*

The 1973 Economic Report of the President devotes an entire chapter (ch. 4) to the economic role of women in the United States. In this chapter, the Report recognizes that economic discrimination against women exists and, by the length and thoroughness of the analysis describing its dimensions and consequences, implies that such discrimination constitutes a serious economic (and social) problem. The Report does not attempt to minimize the extent to which job segregation, earnings differentials, higher unemployment rates exist, and the lack of improvement in each component over the last few decades.

As economists, we are particularly pleased to have the official imprimatur of an Economic Report on the view that discrimination does indeed exist. Some economists have the tendency to minimize the importance of nonpecuniary forces in influencing decisions made within the firm, and have been reluctant to admit the possibility of discrimination unrelated to real or perceived productivity differences. We believe that a proper analysis of discrimination is yet to come; such an analysis will have to fuse elements of economics, sociology, psychology, and history. Employers *do* refuse to hire women for certain occupations. Instead they hire men exclusively and pay them more than they would have to pay women of equal ability. The court records are now full of such cases,[1] but such data will never be explained on the basis of a model which includes in the objective function of the employer only monetary profits. Nor can models which assume that employers' decisions about hiring are based on inborn, unchanging, unexplained "tastes" do justice to the social forces, both internal and external to the firm, which bear on such decisions.

Specifically, it is well known that the average woman college graduate who works full time all year ends up with about the same income as the average male high school dropout. The gross earnings differential works out to be between 35 and 57 percent, depending on the data base used to make the calculation. The Report puts the differential due to discrimination at about 20 percent, but this seems low. In a recent article, Isabel Sawhill reviewed seven econometric studies of male-female earnings patterns. In six of them,[2] the differences which could be attributed to discrimination were above 29 percent and ranged up to 43 percent. The seventh study[3] estimated the difference which might be attributable to discrimination as 12 percent, but arrived at this figure by classifying as nondiscriminatory the differences in the distribution of men and women among detailed occupations. Since

* Professors of economics, University of Maryland. We would like to thank Carolyn Shaw Bell, Nancy Gordon, June O'Neill, Isabel V. Sawhill, and Phyllis Wallace for helpful comments on a previous draft.

[1] For a plethora of case materials see K. M. Davidson, R. B. Ginsburg and H. H. Kay.

[2] See J. Morgan et al., Victor Fuchs, M. S. Cohen, L. E. Suter and H. P. Miller, R. Oaxaca, Sawhill.

[3] See Henry Sanborn.

TABLE 1—EMPLOYMENT BY SEX IN OCCUPATIONS CLASSIFIED BY EXTENT OF WOMEN'S REPRESENTATION IN 1960
(thousands)

Occupations in which women were (in 1960)	1960			1970		
	Women	Men	Total	Women	Men	Total
Underrepresented (0–25%)	2,110	31,231	33,341	3,315	32,436	35,751
Well represented (25–45%)	3,503	6,332	9,835	4,470	7,546	12,016
Overrepresented (45–100%)	15,394	5,863	21,257	20,670	6,930	27,600
Total	21,007	43,426	64,433	28,455	46,912	75,367

employers will typically restrict certain jobs to men and since this restriction is a principal mode of discrimination, a major part of the difference in occupational distribution must be classed as due to discrimination. The fact that women in the past have not trained for or applied for such jobs has been due more to women's realism about the prospects for payoff of such training or applications than to women's voluntary embrace of a benign "role differentiation."

The Report places a great deal of emphasis on the fact that women of a given age have up to now averaged less work experience than men of that age. This emphasis on work experience would be justified if experience for women meant the same as experience for men, but, unfortunately, it does not. Women typically are relegated to jobs in which experience adds little to productivity. Consider the newly hired junior executive and his newly hired secretary. They both may have gone to the same college, got the same grades, and even have majored in the same subject. But, for him, experience will mean learning, increased responsibilities, increased contacts, increased self-realization. In her case, it is likely that the development of work skills will have ceased in six months. The Report documents a strong trend towards increase in labor market attachment in women of child-bearing age; this trend will do little good in decreasing the earnings gap unless occupations are desegregated.

The difference in earning power between men and women is an important contributor to the incidence of poverty and welfare dependency in this country. In 1971, 40 percent of families "with female head" were classified as poor as compared to 7 percent of families "with male heads."[4,5] The wages offered to working women (particularly Black women) frequently provide incomes close to or below the amounts welfare mothers get out of welfare.

With respect to job segregation, the Report indicates (on the basis of an analysis of the proportion of women in 197 occupations) that women tend to be concentrated in "women's occupations," and that there has been only a very small change in the direction of less segregation between 1960 and 1970. We have retabulated the 1960 and 1970 Census data by occupation given in the Report in a way which shows quite graphically the continued occupa-

[4] Some of the families with male head are entirely supported by women, so the difference between 40 and 7 percent understates the effect of low earnings for women on poverty incidence.

[5] The quotation marks in this sentence convey our disagreement with the Census Bureau's use of the term "head." Census cannot mean by it "highest paid worker" since any family including a husband is classed as being "male headed" regardless of his earnings. They must mean "family boss."

tional segregation of women (Table 1). In 1960, 73.3 percent of the women workers were in those occupations in which (in 1960) women were grossly overrepresented; in 1970 the proportion was 72.6 percent. The Report's index of occupational segregation, computed somewhat differently, also changed very little. However, this fixity of the degree of segregation meant a deterioration in the position of women, since in 1970 women constituted a greater proportion of the labor force. In 1960, the women in occupations in which women were overrepresented made up 23.9 percent of the labor force. By 1970, the share of women in these occupations to the total labor force had advanced to 27.4 percent. These occupations were already relatively overcrowded in 1960, and as a result, productivity (and wages) were relatively low.[6] This increase in the relative size of these occupations probably increased the amount of overcrowding and further increased the gap between productivity in "men's" and "women's" occupations. This is corroborated by the decrease between 1956 and 1969 in the ratio of female to male earnings (from 63 percent to 59 percent), a large part of which is due to a relative decline in wages of female clerical workers.

The Report documents the fact that women have generally higher unemployment rates than men and that this differential has been more pronounced in recent years. It ties the worsening trend to the relatively large increase in the labor force participation of women. We agree that the increase in labor force participation rate for women is a part of the cause but not for the reason given in the Report. There emphasis is placed on the fact that a person entering the labor force necessarily has a spell of unemployment while looking for a job. While this is true, the length of such spells and thus the unemployment rate is influenced by the number of job slots for which women are considered eligible as compared with the number of women in the labor force. In our view, the unemployment problem of women has worsened relative to the unemployment problem of men for the same reason wage differentials have increased: because of the segregation of women into "women's" occupations which have become relatively more overcrowded due to the relative increase in the female labor force.

The Report also indicates that quit rates and layoff rates are higher for women than for men. The issue of high turnover among workers is one in which cause and effect are hard to disentangle. The common view is that women have high turnover and enter and leave the labor force more frequently than men because of "their (sic) home responsibilities." In fact, of course, women are consigned for the most part to jobs that have very little interest, opportunity, or pay. Typically, women's jobs are also those for which there is no penalty for high turnover; whether one stays or whether one quits and gets another such job immediately or after an interval, the results in terms of pay and advancement remain much the same. Men who happen to be in this kind of job also have high turnover. Women's relatively high quit rates (2.6 percent per month as opposed to 2.2 percent for men, uncorrected for occupational differences) are seen by some as justifying the exclusion of women from good jobs and by others as an effect of their exclusion. Some may argue that women will have to get over their "lesser attachment to the labor force" before better jobs will open to them, and others may argue that employers will have

[6] The number of men in occupations where women are overrepresented increased also. This is not inconsistent with the view that these occupations are overcrowded. An occupation which a college graduate woman cannot leave because of limited opportunities elsewhere may look relatively good to a male high school dropout.

to open better jobs to women before such improvements in women's turnover can be expected. In fact, these two things will have to occur simultaneously.

The Report gives currency to the recently prominent view that a high turnover rate among women is important in explaining their higher unemployment rates. This view derives, in our opinion, from a misinterpretation of the undoubted fact that all unemployed persons have a history of being separated from a job or entering the labor force. Both turnover *and* the balance of supply and demand for a group's services affect its unemployment rate, but the latter factor is likely to be of far greater importance than the former.[7]

There are, in the Report, a number of instances which we feel reflect a great deal of sociological conservatism. The treatment of paid work outside the home is an example:

> Women work outside the home for the same reason as men. The basic reason is to get the income that can be earned by working. Whether—for either men or women—work is done out of necessity or by choice is a question of definition. If working out of necessity means working in order to sustain biologically necessary conditions, probably a small proportion of all the hours of work done in the United States, by men or women, is necessary. If working out of necessity means working in order to obtain a standard of living which is felt by the worker to be desirable, probably almost all of the work done by both men and women is necessary. [p. 89]

This passage, while clearly reflecting liberal intentions, misses some very important reasons why women (and men) work. They work not only to contribute to the family's funding for goods and services, but for greater personal autonomy in spending, for status inside and outside the family, to occupy themselves in an interesting way, to meet people, to have the excitement of being in a contest for advancement, to reduce the amount of housework they do, and to get away from spending all day with their children. The jobs most women now have tend to fulfill these desires to a less satisfactory extent than the jobs men now have, but they frequently fulfill them better than staying home would.

A woman's work also reduces her own financial uncertainty. A working woman whose husband dies or whose marriage breaks up is in a far better economic position than a similarly bereft housewife in terms of experience, entree, contacts, work habits, and asset ownership. In this day of unstable marriages, a woman who refrains from working during marriage is taking a risky position with her own financial future and that of her children. One year after divorce only 38 percent of ex-husbands are in full compliance with court-ordered child support payments. After five years, the figure drops to 19 percent.[8]

One of the reasons given in the Report for women's greater unemployment, turnover, and lower wages is that, "A wife seldom is free to migrate to wherever her own prospects are best" (p. 99). This statement represents prejudicial past practice and is not necessarily the way things will be in the future. Although casual empiricism is our only source of data on this, it seems to us that the practice of considering only the man's career is far less prevalent than it used to be. The economic problem of couples with specialized job requirements is not really going to be an easy one to solve. One possibility for professional people is geographic mobility early in life, with a postponement of marriage until the person has settled into the slot he or she is

[7] See Bergmann for a more extended discussion of this issue.

[8] See K. Eckhard.

willing to occupy from then on. A second is lessened mobility throughout life, which could not necessarily be to the detriment of the quality of life in this country. A third is the removal of institutional barriers to joint employment. A fourth is the couple's agreement to move to a location where the partner who is weakest in the labor market has the best chance.

Along a similar vein, the Report assumes that when women have children there is no alternative but to drop out of the labor force for a number of months or years. The present authors themselves are exceptions to this "rule." Speaking of those higher income women who do drop out, the Report notes "a considerable sacrifice of earnings" and infers "that these women have therefore placed a very high value on the personal attention they can give their children" (p. 107). Here again we would argue that it is unfruitful to analyze decisions concerning child rearing as being based on unchanging internalized tastes. The appearance of Betty Friedan's book seems to have lowered by quite a lot the "value on the personal attention (women) can give their children." Some women now believe that a lower quantity of personal attention measured in terms of time may increase the quality of that attention to the gain of both children and their parents.

One outstanding omission in the Report is any discussion of changes in the distribution of household tasks between husband and wife, an issue which surely goes to the heart of the women's liberation movement.

The Report asserts there is no practical way to assign a market value to the unpaid work performed at home, subject it to income tax, and thus to tax it comparably with other income and so remove the bias in favor of unpaid work at home. This is quite untrue. The new deduction for paid child care is precisely a move to treat paid and unpaid work similarly. (If domestic work is performed by a family member on an unpaid basis, the value of the service is automatically "deductible," since it is not reported as income.) A deduction for salaries for all paid domestic work by non-family members would seem to be the logical extension. If it is argued that this would favor the rich, then the deduction could be reduced for higher income groups and/or a maximum put on the size of the deduction, and further might be made contingent on all family adults being in the labor force or at school.

In discussing policy with respect to employment discrimination, the Report mentions the many Titles, Acts, and Executive Orders which mandate an end to unfair employment, promotion, and pay practices. It fails to mention that enforcement efforts have been almost nil, despite the fact that very few if any firms, universities, or even government offices are in compliance. What enforcement efforts have been made have raised up loud cries of reverse discrimination.

The present authors are firmly opposed to reverse discrimination and believe it unnecessary and undesirable. Given a fair shake, there is no doubt in our minds that women can make it to full equality with men in the job market. The problem, as we see it, is one of how to implement the transition to "sex-blind" hiring practices, in the presence of prevalent conscious and subconscious discrimination and role casting (on both sides of the "hiring hall"). In this context, there is much to be said for sensibly administered hiring goals of a statistical nature in preference to mere employer statements of good intentions. Naturally, we do not mean a 50 percent ratio across the board. If, for example, women constitute 35 percent of those who by objective criteria qualify for a given rung of an occupation, they should have approximately 35 percent of those jobs, especially in large firms whose very size makes it possible to assume that depar-

tures from the average cannot be explained on purely statistical grounds.

To summarize, the chapter of the President's Economic Report on the economic rôle of women is a creditable coverage of the data, of the problems, and of the issues. It touches almost all the bases one would expect. What would have been welcome and what is missing is a rather more open view of what the problem is, and what the future possibilities are for changing the economic and social role of women. While painting an accurate, reasonably bleak picture of existing reality, the Report tends to underestimate the possibilities and need for social change, and to underemphasize the role of noneconomic forces in having brought about the current situation. Almost absentmindedly, it ignores the transformations in social relations and attitudes, and in economic practices which seem to be occurring and which must accelerate if significant change is to occur. Unfortunately also, the Report suggests very little by way of positive programs.

We must also point out that, however much we are pleased about the attention given to women's problems in the President's Economic Report, we deplore the fact that the Council has apparently reduced its concern in the social area more generally. The issue of unemployment, which by legislative history should be the Council's major concern, is given explicit attention in two passages which together occupy about two pages in the Report. One sentence is as much as we could find devoted to the economic problems of Blacks. Poverty has disappeared as a subject of the President's economists' concern. We can only hope that next year's Report will show that this omission is temporary.

REFERENCES

C. S. Bell, "Age, Sex, Marriage, and Jobs," *Publ. Interest*, Winter 1973, *30*, 76–87.

B. R. Bergmann, "Labor Turnover, Segmentation and Rates of Unemployment: A Simulation-Theoretic Approach," mimeo.; Project on the Economics of Discrimination, Univ. Maryland, Mar. 1973.

M. S. Cohen, "Sex Differences in Compensation," *J. Hum. Resources*, Fall 1971, *6*, 434–47.

K. M. Davidson, R. B. Ginsburg, and H. H. Kay, *Sex-Based Discrimination and the Law*, forthcoming.

K. Eckhard, "Deviance, Visibility and Legal Action: The Duty to Support," *Social Problems*, Spring 1968, *15*, 470–74.

V. Fuchs, "Differences in Hourly Earnings Between Men and Women," *Mon. Lab. Rev.*, May 1971, *94*, 9–15.

J. Morgan et al., *Income and Welfare in the United States*, New York 1962.

R. Oaxaca, "Sex Discrimination in Wages," in O. Ashenfelter and A. Rees, eds., *Discrimination in Labor Markets*, forthcoming.

H. Sanborn, "Pay Differences Between Men and Women," *Ind. Lab. Relat. Rev.*, July 1964, *17*, 534–50.

I. V. Sawhill, "The Economics of Discrimination Against Women: Some New Findings," *J. Hum. Resources*, forthcoming.

L. E. Suter and H. P. Miller, "Income Differences Between Men and Career Women," *Amer. J. Sociol.*, Jan. 1973, *78*, 962–74.

The 1973 Report of the President's Council of Economic Advisers: A Review

By CARL F. CHRIST*

The Economic Report of the President is of necessity a somewhat schizophrenic document. This is so because the Employment Act of 1946 (which called it into being) ordains that it seek to serve two purposes that are not entirely compatible.

On the one hand, it is an apology for the President's economic program, prepared by the President and a Council of Economic Advisers appointed by him for no fixed term. In this role, it seeks to show that the economic policies of the President and his party will be successful (and, if the incumbent president has already been in office for a year or more, have already begun to be successful) in solving the grievous problems created by the foolish mistakes made by the other party when it was last in power.

On the other hand, it is a professional economic report on the American economy and related public policy, prepared by three leading professional economists. These three economic advisers are chosen from among those who have devoted their previous careers (and will in all likelihood devote their subsequent careers) primarily to teaching and/or research in economics. In this role, the Economic Report seeks to present to the Congress and the public an analysis that is competent by the standards of the economics profession, describing and assessing the present state of the American economy, how it came to be the way it is, and how it can and cannot be changed by public policy especially at the federal level.

I believe this schizophrenia is unavoidable, unless the rules are changed either so that the President has no professional economists as advisers, or so that an independent body of professional economists is constituted and charged with issuing reports on the economy and public policy without being responsible to any elected official. Neither of these changes in the rules would be desirable. The first would deprive the President, the Congress, and the public of the competence and discipline offered by professional economic standards. The second would divorce the President's economic advice from the different but equally important discipline of the political arena.

The 1973 Report, like its predecessors from 1949 onwards, consists of three very different components. The first is the "Economic Report of the President" proper, addressed to the Congress and signed by the President. Mr. Nixon's 1973 report, like his 1972 report, is just 5 pages long.

The second is the "Annual Report of the Council of Economic Advisers," addressed to the President and signed by the Council. The 1973 version, by Herbert Stein, chairman, Ezra Solomon, and Marina v.N. Whitman, is 174 pages long, including two appendices. It is in this second part that the schizophrenia is most evident.

The third is a statistical Appendix, con-

* The Johns Hopkins University. I am indebted to my colleagues, Louis Maccini and Jürg Niehans, for valuable comments on an earlier draft, and to Mary Anne Matthews for drawing Figure 1.

sisting this year of 94 tables occupying 113 pages, giving a wide variety of economic data for the United States going as far back as 1929.

I. The "Economic Report of the President" Proper

Mr. Nixon's report begins with a review of the good things about 1972: a $7\frac{1}{2}$ percent rise in real output, a reduction of the inflation rate (measured by the consumer price index) to about 3 percent from about 6 percent in 1969, and a reduction of the unemployment rate to 5.1 percent in December 1972 from 6 percent in December 1971. But then his report seems to try to make the employment gain look better than it really is by saying that employment ". . . rose by 2.3 million people from 1971 to 1972, the largest 1-year increase in 25 years" (p. 3). This is true, but hardly remarkable when one notes that the labor force has been growing every year. A more correct impression would be given by saying that the *percentage employed* rose by 0.7 points (from 95.5 to 96.2 percent of the labor force), the largest one-year increase in 6 years.

Mr. Nixon's report sets forth domestic goals for 1973: "Output and incomes should expand. Both the unemployment rate and the rate of inflation should be reduced further, and realistic confidence must be created that neither need rise again. The prospects for achieving these goals in 1973 are bright—if we behave with reasonable prudence and foresight" (p. 4). And "We must prepare for the end of wage and price controls, . . . " (p. 6). This is reminiscent of Mr. Nixon's statement in January 1969, shortly after taking office, that he would reduce inflation without increasing unemployment and without imposing wage and price controls.

One of the most severe difficulties attending economic policy today is that the public has come to expect government to accomplish things that are impossible. It is unfortunate that Mr. Nixon, by such statements as the foregoing, makes this difficulty worse. The evidence strongly suggests it is not possible for the American economy, structured as it has been since World War II, to achieve simultaneously unemployment rates that remain at $4\frac{3}{4}$ percent or less, and consumer price increases that remain at $2\frac{1}{2}$ percent a year or less, without wage or price controls. In the 25 years since consumer prices leveled off at the end of World War II, this has been achieved in only 4 years: 1952, 1953, 1955, and 1965. See Figure 1. In those same 25 years, the average unemployment rate was 4.8 percent, and the average increase in consumer prices was 2.4 percent a year (these data come from Tables C-24 and C-50 in the Appendix of the Report). Virtually every other industrial country finds the same unwanted guest, inflation, at the banquet of full employment. This is not to say that there is any long-run tradeoff between unemployment and inflation: Chile, for example, has experienced inflation rates as high as 80 percent a year in recent years, with unemployment rates not greatly different from those in the United States. On the other hand, several European countries, with only slightly more inflation than the United States, have had far lower unemployment rates.

If the American public insists on keeping the unemployment rate below 5 percent at all times, and on keeping the consumer price index from rising faster than $2\frac{1}{2}$ percent at any time, with no fundamental changes in the structure of the economy, then wage and price controls will be a necessity. However, no major industrial democratic country has been able to impose effective wage or price controls in the face of substantial inflationary pressure without creating black markets, inefficiency, misallocation, and widespread frustration. And neither the United States nor

any other such country has been willing to tolerate controls and their consequences in peacetime for very long. Hence the evidence strongly suggests that it is not possible for the present American economy simultaneously to keep unemployment below 5 percent, and to keep consumer price increases at or below 2½ percent a year, *even with controls*, because the public will not tolerate controls and their consequences for long. At this juncture, when people believe that government can do things that cannot be done, we may possibly try a system of controls, but we will not maintain it for long.

It is not clear whether there is a feasible change in the structure of the American economy that will make it possible to keep unemployment below 5 percent and inflation at or below 2½ percent at all times.

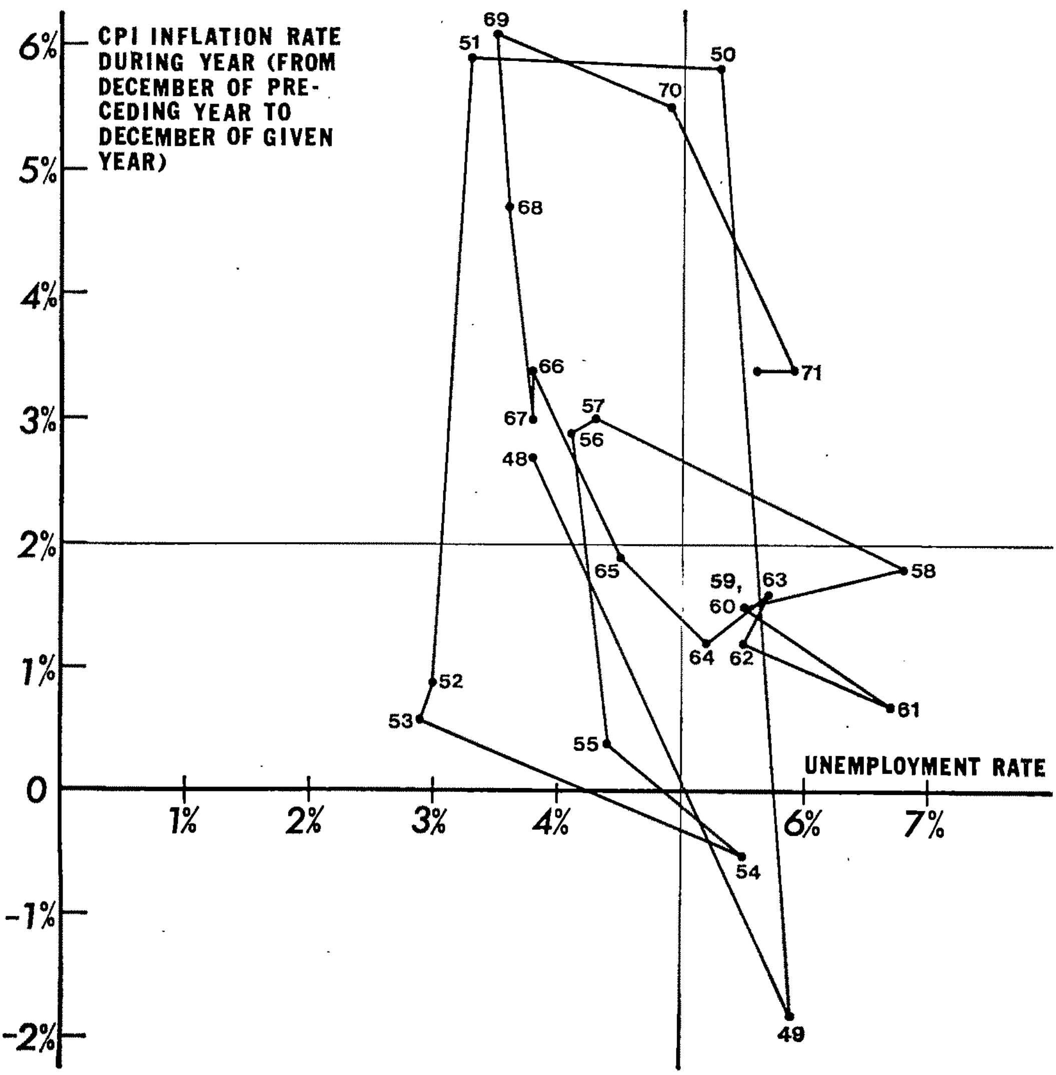


FIGURE 1. INFLATION RATE VS. UNEMPLOYMENT RATE, 1948–72

It has sometimes been proposed (not in the Economic Report) that the largest firms and unions be required to divide into smaller units, with a view to increasing the competitiveness of the economy, but I know of no persuasive evidence that this could accomplish the desired result. Proposals have also been made for better manpower training and job market information.

In my opinion, it would be far better for the President and other national leaders to set forth only goals that are achievable. Then citizens would be less likely to demand government actions that they will regret later (such as the imposition of controls), and more likely to try to solve for themselves problems that government cannot solve for them, and less likely to turn out of office able public officials who have committed no fault except the failure to accomplish the impossible.

Mr. Nixon's report goes on to advocate his well-known ceilings on federal budget expenditures: $250 billion for fiscal 1973, and $268 billion for fiscal 1974. (Even with those ceilings, projected budget deficits for those two fiscal years are $24.8 billion and $12.7 billion.)[1] It urges that government spending proposals be examined by asking whether they are worth their costs, and says "Much Government spending fails this test" (p. 4). This is the Report's only advance indication of the large-scale program cuts that have since been unveiled.

Mr. Nixon's report then outlines his price stabilization proposals for 1973. It summarizes Phase III's easing of wage and price controls except in food, construction, and medical care; asks for a one-year extension of the Economic Stabilization Act (which he opposed in 1970, and said he did not need and would not use); and details several steps already taken to increase food supplies: the suspension of meat import quotas, the relaxation of crop acreage restrictions, the increase in allowed imports of dried milk, the elimination of subsidies on farm exports, and the allowing of cattle grazing on land that has been diverted from crops. (Why were not oil import quotas removed at the same time, or even earlier, instead of later?)

The international economic policy section of Mr. Nixon's report is a plea for the international payments reforms proposed by the Administration in 1972 (described later in the Council's Report), and for support for the Administration's position in the prospective trade negotiations of 1973, laudably couched here in terms of the mutual advantages of free trade and payments.

Mr. Nixon concludes his report with the prediction ". . . that 1973 will be another very good year for the American economy." And that, if we do not exceed the increases in federal expenditures that he has proposed, "It can be a year in which we reduce unemployment and inflation further and enter into a sustained period of strong growth, full employment, and price stability" (p. 7). There go those pernicious promises of the unattainable, again!

II. The "Annual Report of the Council of Economic Advisers"

The Council's Report contains five chapters, entitled "The 1972 Record," "Inflation Control Under the Economic Stabilization Act," "Outlook and Policy," "The Economic Role of Women," and "The International Economic System in Transition." The chapter organization is virtually identical to that in the 1972 report, except that the chapter on women replaces a chapter on microeconomic policy. Appendix A gives supplements to chapters 2, 4, and 5, and Appendix B gives the now

[1] Secretary Shultz announced in May that the faster-than-anticipated rise in money incomes and tax revenues since January had led to a decrease in these projected deficits to $19.8 billion and $5.7 billion, respectively.

traditional listing of all past members of the Council since its beginning in 1946, together with the names of all who served on its staff during the year, and an account of the responsibilities and activities of the Council during the year. Appendix C, containing 94 statistical tables, will be reviewed in Section III. The text has 33 tables and 13 charts.

So many abbreviations are used throughout the text of the report, albeit usually with definitions the first time they appear, that an index might be useful. I offer the following, with the page number where the definition is to be found in each case: *CIEP* 120, *CISC* 60, *CLC* 79, *CPI* 30, *EC* 117, *EEOC* 111, *ESP* 143 (this doesn't mean extra-sensory perception), *FNMA* 49, *GATT* 128, *GNMA* 49, *GNP* 19, *HUD* 84 (not defined), *IMF* 121, *IRS* 146, *JEC* 182, *MFN* 133, *NIA* 19, *NTB* 133, *OECD* 136, *OMB* 181, *Quadriad* 181, *RPD* 45, *SDR* 114, *snake* 117, *TLP* 55, *Troika* 181, *T-2* 181, *T-3* 181, *tunnel* 117, *WPI* 30.

Chapter 1 gives a very informative account of the course of the economy and federal economic policy in 1972. On the whole, it is objective and sets a good standard for material directed to a nonprofessional audience, but on a few occasions statements are made that tell the truth, and nothing but the truth, but not the whole truth, in a manner that might be intended to put the Administration's performance in a favorable light. Here is an example. On page 30 we find this paragraph:

> In the period since the August 1971 policies were instituted there has been a dramatic deceleration in the rate of inflation. This is discussed in detail in Chapter 2; here we focus mainly on price measures applicable to the GNP. For example, the implicit price deflator for GNP, which measures the change in prices of all goods and services included in the gross national product, rose at a rate of 5.2 percent per year in the 3 years ending in the second quarter of 1971, the last full quarter before the policy shift. Over the next 6 quarters, the annual rate of increase was down to 2.7 percent.

This is true, but it creates the incorrect impression that there was an abrupt decline in August 1971, or thereabouts. Actually, the peak rate of increase in the *GNP* deflator, 6.5 percent, was attained in the first quarter and again in the fourth quarter of 1970, and by the second quarter of 1971, before the policy shift, the rate of increase had fallen to 4.4 percent (these are seasonally adjusted annual rates from Table C-4 in the Appendix). And the peak yearly increase in the *CPI* was in 1969. Hence a substantial decline in the rate of inflation had already occurred before the August 1971 policy shift.

Table 13 and Chart 5 point out that the share of interest plus profits before tax (after adjusting to a uniform method of calculating depreciation) in the gross product of nonfinancial corporations declined from 21 percent in 1950 to 17 percent by 1954, fluctuated near 17 percent through 1968, and fell to between 13 and 14 percent in 1970–71–72. The share of interest grew from about 1 percent to about 3 percent over the period, so that the share of profits alone fell from about 20 percent to between 10 and 11 percent. This is broadly consistent with the decline in corporate profits before tax as a fraction of *GNP*, from 13 percent in 1950 to about $7\frac{1}{2}$ percent in 1972. To put this in perspective, note that 1950 had the highest value for the share of corporate profits before tax in *GNP* over the period for which annual national income accounts are available (1929–72), and the figure for 1929 was 10 percent.

Chapter 1 discusses fiscal policy in the usual way in terms of federal receipts and expenditures, and monetary policy in the usual way in terms of interest rates and the changes in the stocks of money, time

deposits, and *RPD*'s (reserves available to support private nonbank deposits). It is useful also to link fiscal and monetary policy together by means of the government budget restraint, which provides that federal expenditures must be financed by some combination of tax revenues, borrowing from the private sector (including state and local governments and the rest of the world), printing new high powered money (which is done by borrowing from the Federal Reserve), and dipping into Treasury stocks of financial assets. It is useful to consolidate the Treasury and Federal Reserve sectors, so that their claims against each other cancel out. Then a Federal Reserve open market purchase or sale is treated just like a Treasury retirement or issue of debt vis-à-vis the private sector. The government budget restraint then implies that the federal deficit during any period must be equal to the following: the increase in the high powered money stock, plus the increase in federal debt held by the private sector, minus the increase in member bank borrowings from the Federal Reserve, minus the increase in *U.S.* reserve assets including gold, minus the increase in Treasury deposits at commercial banks, plus a small balancing item consisting of the increase in all other net claims of the private sector against the consolidated Treasury and Federal Reserve sector. For calendar 1972, the amounts are as follows, in billions of dollars:

deficit		increase in high powered money		increase in federal debt held privately		increase in member bank borrowing
17.4	=	4.3	+	14.6	−	0.9

	increase in U.S. reserve assets		increase in Treasury deposits		increase in other net claims against Federal sector
−	1.0	−	0.1	+	0.5

This means that the federal deficit of \$17.4 billion in calendar 1972 was financed chiefly by borrowing \$14.6 billion from the private sector (thus increasing privately held federal debt by 5.9 percent from \$247.9 billion to \$262.5 billion), partly by issuing \$4.3 billion of new high powered money (thus increasing the high powered money stock by 5.1 percent from \$83.8 billion to \$88.1 billion), and the negative remainder by a \$1.5 billion decrease in other net claims against the federal sector. The fact that reductions in reserve requirements in November released about \$3 billion of reserves made it possible for the money stock to increase by \$20.1 billion, or 8.3 percent, from \$242.8 billion to \$262.9 billion. (These data are from the *Federal Reserve Bulletin* for February 1973, and incorporate the revised figures for money and reserves published therein.) The projected deficits of \$12.3 billion for the six-month period ending in June, 1973, and \$12.7 billion for the fiscal year 1974,[2] presage either tight money if they are financed by borrowing from the private sector, or more inflation if they are financed by the issue of additional high powered money (unless reserve requirements are increased). If none of these alternatives is acceptable, then the deficits must be reduced either by limiting spending even more than proposed by Mr. Nixon, or by increased tax revenues.

Chapter 2, and the appurtenant section of Appendix A, constitute the report on economic stabilization activities that is required by the Economic Stabilization Act as amended. Its treatment of the effectiveness of the Phase I and Phase II controls is more balanced than that of chapter 1. Its main points are that although by August 1971, the rate of inflation had declined from its 1969 peak, and output and employment was expanding,

[2] But recall the revised projections mentioned in fn. 1.

it was uncertain whether the rate of inflation would continue to decline or would again accelerate (pp. 51–52); that if inflation did accelerate again the price increases would not be the direct result of excess demand but would occur because firms and workers preferred raising wages and prices since others had been doing so (pp. 52–53); that controls were necessary on a temporary basis to reduce the risk and the fear that the rate of inflation would rise or not decline further (p. 53); that the controls should be combined with initially expansive but gradually more restrictive fiscal and monetary policy because controls would not in fact be able to hold down prices and wages if there were strong excess demand, and the effort to do so would result in shortages and distortion of production (pp. 53–54); that the controls should be temporary because the Administration expected that if the total package of policies were well managed it would be possible to enter a period of reasonable price stability without controls (p. 54); that the rate of inflation of prices and wages did decline significantly during the controls period (pp. 56 and 58); that the Council believes it is probable that the controls did reduce the rate of inflation, but the magnitude of the reduction is uncertain, and one cannot be sure that the controls had an effect on the rate of inflation (pp. 61–62); that the controls were valuable in any case because they reduced the risk of inflation and reduced expectations of future inflation (pp. 62–63); and "Moreover, what had happened was not merely a suppression of price increases that would burst out if controls were removed" (p. 68).

This chapter's theory of the operation of wage and price controls in 1971–72 appears to be as follows (see Figure 1): in 1966 through 1970, inflationary expectations grew and created inflation without serious excess demand; by 1971 and 1972 there was substantial excess supply, but inflation would have (or might have) continued anyway based on expectations built-up previously; controls reduced the inflation rate by damping expectations about future inflation, without creating shortages or misallocation; controls could then be relaxed in 1973 and dropped later without leading to further inflation or unemployment, because expectations had been restored to normal by the end of 1972. I believe a series of events like this could certainly occur, provided that the control period is long enough and has sufficient excess supply so that inflationary expectations are indeed destroyed, and provided that the levels of inflation and unemployment to which the economy is expected to return after its inflationary binge are simultaneously maintainable. But neither of these provisions has been met in this case.

In the light of events since the publication of the Report in January, two of the statements in this chapter have turned out to be particularly inappropriate. One is on page 63: "By the end of 1972 the American anti-inflation policy had become the marvel of the rest of the world (Table 18) [comparing inflation rates among countries]. Largely because of this change the rest of the world is willing to hold increasing amounts of dollars." Hardly two weeks later, the dollar had been devalued again, and shortly after that attempts by other countries to support the dollar even at the devalued rate were abandoned. The other is on page 143: referring to the 1972 controls, the Report says "However, the temporary system had to be designed to avoid a price upsurge once the controls were removed or relaxed." Experience so far (in May 1973) with the relaxed controls of Phase III suggest that this design was not achieved.

There is a good description of the administrative regulations and procedures

used in implementing the controls. This should prove valuable if later on we need to remember the advantages and disadvantages of the methods adopted. The Report refers readers to the quarterly reports of the Cost of Living Council for more detail. The program cost about $80 million a year during Phase II, and employed about 4,000 persons, as compared with 60,000 during World War II, 15,000 during the Korean War, and about 2,000 during Phase III. The Price Commission received over 10,000 prenotifications of price increases, there were over 3 million inquiries from the public, and about 300 legal actions were undertaken for noncompliance under Phase II.

Prices charged by the federal government were exempted with the understanding that other means would be used to keep these prices consistent with the goals of the stabilization program. One wonders what those other means were, in view of the fact that from 1972 to 1973 the cash-and-carry price of the Economic Report was raised by 33 percent from $1.50 to $2.00, and the postpaid price was raised by 57 percent from $1.50 to $2.35.

The Administration's goals and the Council's forecasts are given in chapter 3, along with a discussion of the Administration's proposals to achieve its goals.

The Employment Act of 1946 calls for "... maximum employment, production, and purchasing power" (Report, p. 71). By common consent, price stability and balance-of-payments equilibrium have been added to that list of desiderata.

The stated goals are that unemployment should be down to the neighborhood of $4\frac{1}{2}$ percent by the end of 1973, and perhaps lower later, though this is not made clear (pp. 73–74), and that the rate of inflation should be down to $2\frac{1}{2}$ percent or less by the end of 1973 (p. 81). I have argued earlier that these two goals are not simultaneously attainable year after year, and that it is a disservice to announce them as if they were attainable.

The forecasts made for 1973 are as follows (pp. 82, 86, and 87). Nominal *GNP* for the year 1973 will rise 10 percent from the 1972 level of $1152 billion to $1267 billion; the chance that this prediction will err by more than 1 percent (i.e., that *GNP* will be less than $1254 billion or more than $1280 billion) is "extremely small." Nominal *GNP* will rise by 9 percent from $1196 billion in the 4th quarter of 1972 to $1303 billion in the 4th quarter of 1973. Real *GNP* for the year 1973 will rise $6\frac{3}{4}$ percent from the 1972 level of $790 billion to $843 billion, in 1958 dollars. The *GNP* deflator for 1973 will rise 3 percent from 145.88 in 1972 to 150.26. The *CPI* will rise $2\frac{1}{2}$ percent or less from 127.3 in December of 1972 to 130.5 or less in December 1973. The seasonally adjusted unemployment rate will decline from 5.1 percent in December 1972 to the neighborhood of $4\frac{1}{2}$ percent by December 1973.

Some of the Council's 1972 forecasts, from pp. 101 and 108 of the 1972 Report, and the actual outcomes, are as follows: Nominal *GNP* was forecast to be $1145 billion, low by $7 billion or 0.6 percent as compared with an actual $1152 billion. Real *GNP* was forecast to be about $784 billion, low by $6 billion or 0.8 percent as compared with an actual $790 billion, in 1958 dollars. The *GNP* deflator was forecast to rise $3\frac{1}{4}$ percent, high as compared with an actual rise of 3.1 percent. The inflation rate was forecast to be down to 2 to 3 percent by the end of 1972; in the 4th quarter the *GNP* deflator rose at an annual rate of 2.7 percent, and the *CPI* rose at an annual rate of 3.5 percent. The unemployment rate was forecast to decline to the neighborhood of 5 percent by the end of 1972; the actual rate in December was 5.1 percent. These forecasts were quite good, except for the year-end inflation rate which was somewhat underpredicted.

I wish I could believe the 1973 forecasts of unemployment and inflation, but for reasons given above, I cannot. Except for the Phase III controls, which are a watered-down version of Phase II, the Report offers no new departures in macroeconomic policy, and no reason to think that we can accomplish now what we have been unable to accomplish before.

Chapter 4, with the appropriate part of Appendix A, gives a highly informative account of the economic role of women. It is difficult to summarize, because there is so much detail. Among the important findings are these. The labor force participation of women is rising, in two respects: the participation of individual women in most age groups is rising as they become older, and furthermore, the participation of each cohort of women at a given age exceeds that of the previous cohort at the same age (Chart 9). The more education a woman has, the more likely she is to work, except that if she has children under 6 years old, she is unlikely to work regardless of her education (p. 95). Women change jobs less frequently than men, but spend fewer years at each job than men, because they leave the labor force more than men do, the most common reason being pregnancy (pp. 95–96). Black women have a higher labor force participation rate than white. Black married women earn about three-fourths as much as Black married men, while for whites the figure is about one-half, considering in each case only those who work full time the year around (p. 96). The unemployment rate is higher for women than for men, but the average duration of unemployment is less for women than for men; the apparent anomaly is explained by the fact that labor force entrants are unemployed for shorter periods than other persons, and women enter (and leave) the labor force more than men do (pp. 97–8). The differential in unemployment rates in favor of men has widened in the past decade, but part of this is due to a change in statistical concepts in 1967 (pp. 99–100). There has been a small reduction in the difference between the occupational distributions of men and women between 1960 and 1970 (pp. 100–03). Median annual earnings of women are about 40 percent of those of men; adjustments for differences in hours worked per week, weeks worked per year, age, education, experience, and continuity at work increase the ratio to 80 percent (pp. 103–06). The chapter concludes with a discussion of the problems of a woman who is the head of her household; difficulties in making the income tax treatment of women fair; the pros and cons of furnishing child care at public expense; and government programs to promote the economic equality of women.

Chapter 5 deals with *U.S.* international economic relations, and problems of the international trade and payments systems. This chapter begins with a brief account of *U.S.* balance-of-payments developments in 1972. The rest is a longer informative account of the recent history and current state of the international payments mechanism, and of the reform proposal put forward by the United States in 1972. Part of Appendix A is devoted to more detail on that proposal.

The Council says properly that an international monetary system must meet three requirements if it is to serve the needs of world trade and investment: it must be *market-oriented*, *multilateral*, and *stable*. And to be so, it must provide for the *adjustment* of payments imbalances, and for adequate international *liquidity*, and must create and maintain *confidence* in its continued viability and in the value of international reserves (pp. 120–21, Council's italics).

There are several references (for example, pp. 122, 131, 162) to the failure of the Bretton Woods system (with the dollar

pegged to gold, and other currencies pegged to the dollar) to provide adequate growth in the international reserves, and to the creation of *SDR*s ("paper gold") as a means of solving this problem. The Council does not take note of the fact that an alternative way of solving the liquidity problem would be to permit the money price of gold to rise. Since 1933, just before the United States increased the gold price to $35 an ounce, the official price of gold has risen by a factor of only 2.0 (from $20.67 to $42.22 per ounce), while the consumer price index, the wholesale price index, and the *GNP* deflator have risen by factors of 3.4, 3.5, and 4.0, respectively. At the time of this writing, the free market gold price stands at $91 per ounce, about 4.4 times the official 1933 level of $20.67. While it is by no means clear that $70 to $90 would be an equilibrium dollar price for gold if there were no international agreement to fix its price in transactions among central banks, neither is it clear that central banks and the *IMF* will be able to succeed in displacing gold as the main international currency in favor of paper currency such as *SDR*s. That depends upon the demand for gold among the world's asset-holders, including central banks, which will depend in part upon whether the new international paper currency is managed in a way that maintains its long-run values as compared with gold.

As for adjustment of payments imbalances, it is clear, from the fact of several dozen devaluations (and a few revaluations) since the Bretton Woods agreement was instituted after World War II, that exchange rates do require to be adjusted from time to time, whether the system is nominally one of fixed rates or flexible rates. It is not clear whether the day-to-day burdens of hedging against small changes in flexible rates are greater or less than the sum of the intermittent burdens of hedging against occasional large changes in fixed rates, plus the burdens of maintaining domestic policies that are undesirable for their own sake but are followed for the sake of defending a fixed rate in the face of market pressure against it. I lean toward flexible rates, except as between regions that have elected to have a common monetary and fiscal policy, such as the European Community countries have been in the process of deciding about for the past several years, and such as the United States in effect adopted in 1789. In my view, the fact that the Council and other high government officials can discuss exchange rate adjustment mechanisms on their merits is an improvement over the situation of several years ago.

The Council urges "innovations" which will permit to the United States the same margin of adjustment of exchange rates as to other countries, as contrasted with the Smithsonian agreement under which the dollar could not deviate from other currencies by more than $2\frac{1}{4}$ percent, while others could deviate from each other by as much as $4\frac{1}{2}$ percent (pp. 126–27). Presumably the innovations required would consist of pegging all currencies to gold, or to the *SDR*, or to each other as in the European Community, or not at all, rather than to any key currency such as the dollar was.

The Council properly takes a dim view of selective adjustment measures, such as tariffs, import surcharges, and quotas, and properly an especially dim view of quotas, since they permit the least amount of market response to changes in conditions. I would like to note that a sliding tariff, continuously equated to the difference between the domestic and world prices of a commodity, is nearly as bad as a quota, and should be discouraged by international agreement just as should quotas.

The Council supports the intent of the *U.S.* reform proposal, to devise means to press surplus countries to revalue, as a

counterpart to the natural pressure to devalue that deficit countries feel. (But where was *U.S.* zeal on this point in 1949 when *U.S.* reserves reached $26 billion, over half of the world's total?) At present this is a less pressing problem than it seems, because the need now appears to be more for an increase than a decrease in real international balances; hence every time a Japan or a Germany revalues in terms of gold and the *SDR*, the effect on *total* real reserves is pernicious because the total declines rather than increasing; it would be better for the required exchange rate adjustments to be made by the devaluation of overvalued currencies instead.

The heart of the *U.S.* proposal is that for each currency there should be established a theoretical "base level" of reserves, and also below that a "lower point," and above it an "outer point." Whenever a country's reserves fell below its lower point for a period of time, that country would be expected to devalue, and would suffer sanctions if it did not, unless the *IMF* decided (presumably in response to an appeal) that sanctions were not warranted. Similarly, whenever a country's reserves rose above its outer point for a period of time, that country would be expected to revalue, and would suffer sanctions if it did not, again unless the *IMF* decided that sanctions were not warranted (see pp. 124–25, 166–67). If the system were designed in such a way that the imposition of sanctions were highly uncertain in each case until the *IMF* had made its decision, it would be similar to both the Bretton Woods and Smithsonian agreements, except for its added pressure on surplus countries. However, if it were rigorously administered, so that countries passing their "lower" or "outer" points were all but certain to suffer sanctions and hence to be forced to devalue or revalue, with virtually no exceptions, then it might render international capital flows much more violent and severe than they are at present. For if everyone could rely nearly absolutely on, say, a devaluation of the dollar whenever *U.S.* reserves fell below the *U.S.* lower point, then whenever *U.S.* reserves approached the lower point there would be very heavy flows out of the dollar and into other currencies, thus assuring the devaluation of the dollar.[3] Speculation against a weak currency can only be stemmed if the speculators are not sure they will make a profit. Hence the advantages of the *U.S.* reform proposal would seem to be greater, the less vigorously it is applied.

III. Appendix C, Statistical Tables

For many readers, the 94 statistical tables in Appendix C are the most useful part of the Report. They constitute the cheapest, handiest, and most promptly available compendium of economic statistics on the *U.S.* economy that I know of. For many years I have acquired two copies of the Report, and kept one at home and the other at my office, so as to have a copy always at hand when I want to look up an economic datum. The tables cover national accounts; population and labor statistics; production indexes and other business activity measures; prices; money and credit; government finance; coporate finance; agriculture; and international economics.

Do not throw out the 1972 Report if you make extensive use of these statistical tables. This year, for the first time, the tables do not give the complete annual time-series since 1929, but give instead only the figures for 1929, 1933, 1939, and then every year since 1939. The purpose is to leave room at the foot of the page to show monthly or quarterly data for the most recent few years. Hence you will need to keep the 1972 Report alongside the

[3] I owe this point to Jürg Niehans.

current one in order to have complete time-series since 1929.

There are four new tables this year: C-4 gives several different versions of the *GNP* deflator, using different weights and concepts, and also shows percentage changes, year-to-year and quarter-to-quarter. Table C-11 gives nominal and real gross product originating in nonfinancial corporations, and nominal costs per unit of real gross product. Table C-33 gives percentage changes in output per manhour, unit labor costs, and the related data. Table C-47 gives seasonally adjusted consumer price indexes. The total number of tables is the same as last year; six of last year's tables (28–33) have been condensed into two of this year's (C-30 and 31).

IV. Concluding Remarks

I looked in vain for discussion of several economic problems that are generally regarded as among the most important facing the nation today. One is the welfare program, and alternatives to it. Since we have so far been unable to keep the unemployment rate below 5 percent, let alone $4\frac{1}{2}$ or 4 percent, and since we are unlikely to be able to do so in the future either, there will continue to be a function to be served by the unemployment compensation and welfare systems. The present welfare system has been widely criticized for many of its features, from its rapidly rising costs to the degree of effectiveness with which it meets the problems of poverty. Proposals for reform have been offered, and experiments have been conducted to learn about their effects. An analysis of these proposals, done as carefully as the Council's Report, would be a substantial contribution to public understanding and to an eventual solution of this vexing socioeconomic problem.

A second neglected issue is that of federal tax reform. The present federal tax system is unfair to the extent that it taxes unequally persons having the same incomes. The top-bracket personal income tax rate is now 70 percent. But few people having high income from property rather than from work pay this rate, or anything like it, for the present tax law contains many provisions permitting special favorable treatment of certain kinds of property income. Hence it is commonplace for a taxpayer who receives wages, salaries, dividends, and/or federal or private interest to pay a far higher tax than another taxpayer who receives an equal or even larger income from capital gains, oil and gas property, real estate development, certain foreign commerce, or interest on state and municipal bonds. Indeed, taxpayers who receive dividends are taxed twice, once via the corporate income tax, and again via the personal income tax. There are also important issues with respect to the estate tax and the gift tax. Mr. Nixon stated during his 1972 election campaign that he would present a tax reform package in January, and the Council's Report would have been an appropriate vehicle for it. (Secretary Shultz made a modest beginning when he presented the Administration's mild tax reform and tax streamlining proposals in May.)

Other important economic policy problems receiving little or no attention in the Report are environmental preservation, energy, health and health care, and the economic role and status of minority groups. To be sure, last year's Report discussed pollution control, energy problems, and health care. Nevertheless, these issues are no less important now than they were a year ago, to say the least, and since few people remember or go back and re-read old Economic Reports of the President, each new one ought to deal with the economic problems that are currently, or soon will be, of high importance.

The 1973 Report of the President's Council of Economic Advisers: Whistling in the Dark

By DAVID I. MEISELMAN*

The 1973 Economic Report is full of surprises, most of them disappointing. Among them is the apparent certainty of views, just before the storm, that inflation was tapering off and would continue to do so. For professional economists inured to the hazards of the forecasting trade, the Report contains several surprises of greater intellectual and academic interest. The first is the highly sophisticated analysis of the economic role of women. The second is one of the better briefs for the use of price and wage controls to stabilize prices, which is included in the Council's Report as a separate chapter. I share the judgment of Reuben Kessel, Edmund Phelps, Armen Alchian, and others who reviewed the Council's case for controls in the 1972 Report that not even the best case is good enough. Another surprise to many readers will be the strength and persistence of the Report's Keynesian bias, only partly explained by the Employment Act of 1946. The Report's discussion of government expenditures is not surprising but it is both noteworthy and by far the most rigorous and constructive part of the entire Report (see pp. 76–78). Achieving the required reduction of government expenditures and taxes implied by the discussion would truly be surprising.

The most disappointing aspects of the Report are the weaknesses and inconsistencies of its underlying analysis and its clever but faulty defense of and apologia for Phase I and II wage and price controls. Admittedly, the Nixon Administration did have some bad luck because of poor weather and the like. With the safety and wisdom of hindsight, it is now abundantly clear that the economy was excessively stimulated in 1972. As usual, the expansion of aggregate demand affected output and employment first and prices only later. No doubt, the improved economic performance of 1972 plus the control program, or the appearance of one, helped deprive the Democrats of important issues prior to and during the 1972 presidential campaign. By 1973, we had overshot the mark. Inflation had been ratcheted up to a new and higher trend. As before, slowing the accelerated inflation is bound to be costly. The goal of stable prices is more distant and more difficult to attain.

Errors in forecasting inflation are inevitable given the limitations of our understanding of how short-period changes in nominal *GNP* are divided between changes in prices and changes in output. Errors of prediction resulting from unavoidable ignorance and the poor state of the art may explain poor forecasts and help to excuse some of them. They need not lead to biased results for the latter require a corresponding inflationary bias on the part of policy makers.

* Professor of economics, Virginia Polytechnic Institute and State University. I received helpful comments from Gottfried Haberler, Wilson Schmidt, Deane Carson, and William Rule.

In my judgment, the 1972 overstimulation is much more serious than still another forecasting error because it is added evidence regarding the strong and inherently inflationary bias of virtually all modern governments. The bias is the result of a combination of factors which have increased the incentives and the technical means for governments to cause inflation at the same time that the constraints on their abilities to do so have been severely weakened. The bias increases in election years. For those in or close to the seats of state power, the costs of paying the piper fall sharply after a successful election day. The myopia of voters, including those with Ph.D.'s in economics, makes the game possible. Are we destined to add to our list a new cycle with a periodicity of four years?

There were impressive economic gains in 1972 in higher output and employment and lower rates of inflation and unemployment. Readers will find the good economic performance of 1972 attributed to the New Economic Policy of August 15, 1971. It combined controls aimed at curbing inflation with an expansionary fiscal policy in the form of both tax reductions and "... boosting Government expenditures in the first half of 1972, mainly by bringing forward expenditures that would have been made later" (p. 4). Although the 8.3 percent growth in money during 1972 was the highest of any year in the post-World War II period, monetary policy is taken to have been essentially passive and accommodating both because of the general stability of interest rates and the fact that nominal *GNP* increased somewhat faster than money did. Thus, a literal reading of the Report should provide much comfort to orthodox Keynesians and Federal Reserve traditionalists by allaying their fears that the Nixon Administration had become captives of the Monetarists, or had even seriously listened to them during 1972.

Consistent with President Nixon's earlier announcement that he had become a Keynesian, the Administration essentially followed the policy of its political opponents by relying on strong stimulation of aggregate demand to lower unemployment and Phase I and II controls to lower inflation. For a short time, it seemed to work.

The fact that much of the Report was soon overwhelmed by events palls the usual self-congratulatory tone of what has become part of the annual ritual of the Economic Report. For example, on the eve of the sharpest spurt of inflation since at least the Korean War and less than a month before the 10 percent devaluation of February 12, the Annual Report of the Council of Economic Advisers noted, "By the end of 1972 the American anti-inflation policy had become the marvel of the rest of the world. . . . Largely because of this change the rest of the world is willing to hold increasing amounts of dollars" (p. 63). The Council's forecast, which did so well in 1972, quickly became a shambles early in 1973.[1]

The poor forecasting performance of the Report cannot be attributed to analytical innovations that did not pan out. In my judgment, one of the deficiences of the Report is that it conforms mainly to the format and the content of the Council's analyses of yesteryear. Its framework is the simple Keynesian model. The Report largely ignores how thoroughly the model

[1] The embarrassment has been both bipartisan and virtually universal. For example, sharing the optimism of the Republican president, Walter Heller crystal-balled the 1973 *GNP* deflator as rising only slightly to 3-1/3 percent from the 3 percent rate of 1972 as part of a strong expansion ". . . without fatal flaws of malfunctions" (*Wall Street Journal*, Jan. 11, 1973). The large file of *GNP* forecasts I collect shows that not a single published forecast came close to calling the shots on the inflationary outburst in the first half of 1973 and the 8 percent increase in real *GNP* in the first quarter.

has been modified, both by a vast body of scientific evidence and analysis and by repeated failures of past public policies that depended upon it. The Report also generally overlooks the rich and important advances in recent years in establishing the close connections between micro- and macro-economics. In particular, it underplays the central roles of relative prices, lags, and adjustment processes in short-period fluctuations in helping to explain labor market phenomena such as employment and unemployment. The Report's annual review of the 1972 labor market is devoid of analysis and is largely a verbal repetition of the numerical contents of several standard tables in the *Monthly Labor Review* describing employment, unemployment and labor force participation. The Report would have been greatly strengthened if the highly sophisticated and informative analysis of the economic role of women had been applied to the remainder of the population. Reverse discrimination is also costly.

The review and outlook chapters have typically been the central features of Council reports. The 1973 Report follows this tradition. The review chapter has five sections plus a short postscript on fourth quarter 1972 developments. The five sections are devoted to "Demand and Output," "The Labor Market," "Prices and Costs," "Fiscal Policy" and "Financial Policy and Financial Markets."

Following another tradition, the underlying macro theory is never clearly or explicitly spelled out, but reading the entire chapter does suggest the general outlines of one. Nominal *GNP* is taken to be mainly determined by private expenditures for business fixed investment and for housing construction, both of which have responded to tax incentives, financial variables and autonomous factors, all expansionary in 1972. Nominal consumer spending (i.e., personal consumption expenditures) is induced by the resulting general improvement in the economy. Reference is made to the fact that consumers "made extensive use of credit to finance their purchases, particularly of durable goods" (p. 23), but it is not clear whether the quantity of credit (no cost is mentioned) is the result of the increase in consumer outlays or a cause. Government expenditures are considered in their full employment budget context. The chapter on the 1972 record has no accounting for foreign trade, which is treated in a separate chapter essentially unrelated to the *GNP* accounts.

Aggregate nominal demand is also related to fiscal policy, mainly as evaluated by high employment federal government expenditures and receipts. The net stimulus is taken to be the change in the high-employment deficit. In this view, changes in nominal high employment expenditures and receipts have equal and opposite effects, modified slightly by a permanent-income hypothesis view of temporary changes in tax cash flow, such as the 1972 overwithholding.[2] Only the federal budget is considered.

The emphasis on the two major components of gross private domestic investment (the change in inventories is omitted) overlooks the facts that the empirical association of total investment and *GNP* is poor indeed, and that there are important differences in the timing of the business investment and housing expenditure components. Despite continued emphasis by most business forecasters in government and elsewhere on the role of business investment expenditures in determining

[2] In the Economic Report of the President, President Nixon's statement about the federal budget emphasizes both the inflationary and allocative effects of federal government expenditures and so is not entirely in agreement with the Council's stated view. Apparently, in his view, a large budget per se is inflationary.

GNP, it turns out that these outlays typically follow rather than lead the business cycle. Indeed, business expenditures for new plant and equipment are a lagging indicator! By contrast, expenditures for residential structures lead the cycle.[3]

Collateral emphasis on the size of the federal deficit in controlling *GNP* ignores the empirical studies of Friedman-Meiselman, Andersen-Jordan, and others which bring into serious question the keystone of the fiscalist analysis and the efficacy of policy proposals stemming from it. The Andersen-Jordan results are consistent with President Nixon's assertion cited above about the expansionary impact of the sharply higher level of federal government spending in 1972, but recall that this works only in the very short run. When money is held constant, an increase in full employment expenditures temporarily increases *GNP*, but later causes *GNP* to decline. Within a year the net effect is essentially zero. The expenditures merely cause an intertemporal shift in the total.

Even within the analytical and empirical framework that attributes an important role to government expenditures and taxes, it would seem that the general omission of state and local budgets is a serious deficiency. Analyzing government on a consolidated basis, following the practice used in the consolidated government receipts and expenditures account of the national income accounts, is highly desirable, indeed, necessary. State and local government expenditures are huge indeed, and have been growing even more rapidly than the federal government's. By 1972 state and local government expenditures on the national income accounts basis were $162.9 billion, or almost 80 percent of the $209 billion for federal expenditures net of grants to state and local governments. Similarly, focusing solely on the federal deficit or surplus can also be highly misleading. In 1972 the federal deficit on the national income accounts basis (net of grants to state and local government) was $18.5 billion. The federal high employment budget deficit was $6.0 billion. In 1972 state and local governments recorded a *surplus* of $12.3 billion on a national income accounts basis. We are left to wonder whether government, on net, was exercising restraint or providing stimulus during the year, especially in the fourth quarter of 1972 when revenue sharing sharply increased both the deficit of the federal government and the total surplus of state and local governments.

Monetary policy appears to be evaluated by two criteria. The first, and more traditional, indicator of monetary policy is interest rates and credit market conditions. The rate of interest is essentially taken to be determined by the demand for and supply of cash, with resulting interest rates modified somewhat by expectations of future inflation and interest rates. "Accommodating" is never defined, but whatever it means, monetary policy in 1972 is viewed as having been accommodating because of the general stability of nominal interest rates. The source of the expansion was elsewhere. The Federal Reserve can claim credit only for not having pulled on the string.

There seems to be a second indicator of monetary policy which is more novel, but no more apt. Other evidence for the Fed's having taken an accommodating rather than active role is presented by citing the fact that the money supply increase of 8.2 percent (later revised to 8.3 percent) was less than the expansion of 10.9 percent in nominal *GNP*. This is a position which has been publicly shared by Arthur Burns, Chairman of the Federal Reserve Board. In other words, the fact that income velocity rose slightly during the year, as it typically does during business cycle ex-

[3] See *Business Conditions Digest*.

pansions, is interpreted as support for the contention that monetary policy was essentially passive in 1972.

I also find it difficult to understand how velocity can be used as an indicator of monetary policy. The Fed may directly control the stock of money, but surely it is the public's demand for cash that determines the ratio of money to *GNP* or the response of *GNP* to a given change in money. To suggest that a central bank can treat velocity as subject to its direct control or that it can effectively treat velocity as a target or indicator of monetary policy is similar to but no more valid than the view that it can control real balances.

The novel proposition raises some interesting questions. Consistent with this view would it be correct to infer that the central bank was merely being accommodating in a mythical country which had experienced a doubling of money accompanied by a 120 percent increase in nominal *GNP* or prices? Would a 200 percent increase in prices associated with *GNP* growing still faster than money have been more accommodating or less? Alternatively if, as in the early 1930's, velocity falls when *GNP* declines faster than a central bank has contracted money, does this imply an expansionary monetary policy? What of constant velocity?

Although the Report cites "remarkably mild increases in short-term interest rates," the record is that short-term rates did, in fact, rise sharply and almost continuously throughout the year. Early in 1972, short-term rates had been falling since mid-1971. The low point was reached in February 1972, when 3-month Treasury bills yielded 3.18 percent on average. By December 1972, the average new issue yield had risen above 5 percent. (By March 1973, bill yields were above 6 percent.)

However, the important point is not that short-term rates and income velocity rose, but that there was an 8.3 percent increase in M_1 during 1972. (M_2 increased 10.8 percent.) The monetary base, which is clearly under the direct control of the Fed, also increased by the same 8.3 percent. The base multiplier remained virtually constant throughout the year. Money increased so rapidly precisely because the Fed increased the monetary base. In my judgment, the Federal Reserve bears a large share of the responsibility for the 1973 acceleration of inflation. They blew it again.

The discussions of monetary policy ("Financial Policy and Financial Markets") and of fiscal policy are unrelated, reflecting the lack of unity in the Report as a whole. There is no explicit consideration given to any possible relationship or tradeoff between fiscal and monetary policy, the impact of fiscal policy on interest rates and financial markets and perhaps thereby on resource allocation, the balance of payments, and the like. No mention is made of debt management, a commendable move towards its deemphasis.

The outlook and policy chapter has several guides to the use of fiscal and monetary policies for reaching the potential growth rate of the economy and reasonable price stability by means of a steady increase in nominal *GNP*. For fiscal policy the Report recommends the general rule of achieving and maintaining balance in the full employment budget of the federal government:

> . . . desired . . . steadiness of the rate of increase of money GNP would be assisted by keeping the budget in a position of balance at full employment, unless there later appears clear and strong evidence that developments in the private sector call for greater fiscal stimulus or restraint. Constancy in the relation between full-employment receipts and expenditures provides approximate constancy in the Federal contribution to economic expansion. The constancy of the contribution is only approximate

> because the impact of the budget on the economy depends on its composition and total size as well as on the size of the deficit or surplus. However, constancy of the balance at full employment is the best single guide to a budget policy that neither pushes the economy above its desired growth rate nor holds the economy below it. In any case, the rule that expenditures should not exceed the revenues that would be collected at full employment is essential to bring home the requirement that Government should not spend money for things it is unwilling to ask citizens to pay for. The circumstances of 1973 are appropriate for getting on to that track. [p. 74–75]

The statement does not agree with the Report's companion assertion:

> . . . experience has shown that the proper conduct of macroeconomic policy may sometimes call for a prompt and effective shift in the overall balance between the flow of Federal receipts and expenditures. Temporary and prompt changes in tax rates, which do not alter the basic structure of taxes, may provide an efficient way of accomplishing such required shifts. [p. 75–76]

You obviously can't have it both ways. Moreover, evidence both from the imposition of the 10 percent surtax in 1968 and its removal in two steps in 1969 and 1970 as well as from large numbers of independently replicated statistical tests indicate the weakness of this and related fine tuning tax measures.

For monetary policy, the Report proposes monetary growth at a steady but unspecified rate. This recommendation is inconsistent with the Report's earlier discussion of monetary policy described above. No mention is made of interest rates or velocity as targets or indicators of monetary policy as the Report states:

> A gradual slowing down of the expansion of money GNP to a steady rate consistent with the long-run potential growth rate of the economy and reasonable price stability is also an approximate goal for monetary policy. This is likely to require a slower increase in the supply of money and credit than was proper when the main objective was to encourage the quickened economic expansion in an environment of substantial unused resources. [p. 75]

The parts of the Report devoted to inflation probably received the greatest care and have attracted the most attention. Students of inflation will find many theories of short-period changes in the price level. A variety of theories is understandable in view of the poorly developed state of economics in understanding short-period changes in the price level. Even so, some should be ruled out because of errors of analysis or because they have been contradicted by the evidence. I find that one of the two leading theories in the Report is simply wrong, and the other highly questionable.

One approach believes that the price level is determined by the costs (expenses?) of producing the goods entering the price index. This view confounds relative and absolute prices as well as *ex ante* and *ex post* values. The review section, "Prices and Costs," attributes the 1972 slowing of inflation to a ". . . marked deceleration in the rise in unit costs. For labor costs, this was a result of two factors: a slower rise in hourly compensation and an acceleration of the rise in output per man-hour. . ." (p. 31). It is not entirely clear whether the Report is presenting an identity or a behavioral relationship. National income is the factor expense of the *GNP*; the price level enters both sides of the accounting identity. Explaining one deflator by the other plus the change in productivity is essentially a tautology. Double entry bookkeeping being what it is, it is difficult to see how the result could be very different, especially in view of the stability of factor shares.

Another approach emphasizes the im-

portance of price expectations in contributing to the inflation itself. The general proposition has some merit, but not in the context found in the Report or in most current discussion. It should not be necessary to do so, but it bears repeating that inflation is a change in the value of money, not a change in the average of selected prices.[4] Moreover, changes in the price level are an important part of the general equilibrium process whereby the demand for money becomes equal to the supply of money. This is why there is essential agreement among all economists that for given real factors, including preferences, a once-for-all change in nominal money leads to a proportional change in the price level *in the long run.* For given real factors, velocity is constant. Prices adjust to change the value of real balances in order to reestablish the equilibrium where desired and actual real balances are equal. Because actual and anticipated changes in prices affect the perceived cost of holding cash, they must be included in the real factors determining velocity. Although the short-run dynamics are uncertain, the economy is nevertheless driven to the long-run norm.

In the short-period changes of the business cycle, links between money and prices are clearly looser, in part because income and other real variables change, in part because of lags in adjustment, and the like. These are some of the reasons the short-run analysis of price levels is so much more difficult and uncertain. Yet, a decline in the rate of expected inflation can itself lower velocity, moderating the inflationary potential of a given money stock. If, indeed, we knew how to alter inflation expectations, which we surely do not, we could have some impact on both velocity and interest rates. But if we really knew the effect on velocity we could more easily achieve the same goal by appropriately altering the money stock.

It is worth noting that in 1972 both velocity and short-term interest rates increased. If price expectations had been altered as substantially as the Report claims, how can we explain why both velocity and short-term rates moved in the wrong direction? Obviously, either price expectations did not really change very much and/or the impact of the altered anticipations was swamped by other factors. In any event, it suggests that the shift in expectations played a minor and uncertain role in the events of the year.

Comparing the discussions of price and wage controls in the 1972 and 1973 Reports, I found that the 1973 Report was unable to provide any additional economic arguments for them, except perhaps, at the time the 1973 Report was written, the 1972 deceleration of inflation gave the program some appearance of success. When reviews of the 1972 Report were written for this and other scholarly journals the Phase II controls program was already in place. I believe the critiques of Alchian, Kessel, and Phelps, with whose views I concur, read even better than when they were first written in the spring of 1972. Neither the 1973 Report nor the events subsequent to these reviews seriously challenge their contentions that the controls were, at the very least, disruptive, wasteful, discriminatory, arbitrary, costly —and ineffective.

Because the 1973 surge of inflation followed so quickly on the heels of the relaxation of Phase II controls, it will be difficult for the public not to draw incorrect *post hoc propter hoc* inferences that controls had been effective. Indeed, after the President and his economic advisers and spokesmen in and out of the Council of Economic Advisers had for so long attrib-

[4] For an alternative view, see the analysis of what I termed "The peanut theory of prices" (p. 745), in my 1968 comment.

uted so many good results to the controls, citizens are justifiably puzzled why the reluctance to go still further with even stronger measures. As the 1973 Report said, ". . . nothing seems more obvious than that the imposition of controls should curb inflation" (p. 63). The long and dismal historical record that price and wage controls can, at most, only appear to curb inflation temporarily appears to have been subordinated to the requirements that the Administration do something, even if the actions were detrimental to both economic efficiency and individual liberty.

The added defense, ". . . the controls had a symbolic value as evidence of the Government's determination to take whatever steps might be necessary to check inflation" (p. 63), may help to explain the symbolic acts of the executive branch of the federal government. There has clearly been a clamor for action, and controls give the appearance of action. In a democracy it is difficult, indeed, for any political leader this side of heaven to resist such demands, especially if, in the process, more power and discretion flow to him. I would like to believe the chain of events was not inevitable—that the only options were not between weak and tough controls, especially since I am convinced that there is little on the horizon to alter the worldwide inflationary bias.

Economic policy has suffered from a lack of leadership and candor. It is not too much to ask that the President whose leadership in foreign affairs made possible the first fundamental changes in Amercan foreign policy in a generation, who led the successful campaign to eliminate conscription, and who cut the dollar loose from the incubus of gold and fixed exchange rates, also exercise some of the same qualities of leadership at home.

REFERENCES

A. Alchian, "Review of the Council of Economic Advisers' 1972 Report, A Comment," *J. Money, Credit, Banking,* Aug. 1972, *4,* 704–12.

R. Kessel, "The 1972 Report of the President's Council of Economic Advisers: Economics and Government," *Amer. Econ. Rev.,* Sept. 1972, *62,* 527–33.

D. Meiselman, "Comment," *J. Polit. Econ.,* July/Aug. 1968, Part II, *76,* 743–50.

E. S. Phelps, "The 1972 Report of the President's Council of Economic Advisers: Economics and Government," *Amer. Econ. Rev.,* Sept. 1972, *62,* 533–39.

U.S. Department of Commerce, *Business Conditions Digest,* various issues, Washington.

The 1973 Report of the President's Council of Economic Advisers: A Critique

By MICHAEL PARKIN*

The record of the American economy in 1972 looked impressive. The major preoccupation, the rate of inflation, was reduced (as measured by the Consumer Price Index) from almost 5 percent in 1971 to a little over 3 percent by the end of 1972. This was achieved at the same time as the unemployment rate was being cut back from its 6 percent peak in 1971 to end 1972 at 5.1 percent, and as the real growth rate (as measured by constant price *GNP*) was being stepped up from its 1971 level of 2.7 percent to 6.5 percent. The prospects for 1973, as seen by the Council of Economic Advisers (*CEA*) in the Report point to an equally impressive performance. The inflation rate is seen as falling further to the region of $2\frac{1}{2}$ percent, unemployment as falling further to the $4\frac{1}{2}$ percent zone, and constant price *GNP* as growing by even more than in 1972 at about $6\frac{3}{4}$ percent.

The view of the Report is that 1972's successes were due to a combination of the skillful use of fiscal and monetary policy and to the wage-price controls associated with the Economic Stabilization Act amendments initiated on August 15, 1971. "Fiscal policy was deliberately expansionary in 1972" (p. 40) with stimulation coming from both spending increases and tax reductions. The money stock (narrow definition) grew at 8.9 percent on the year and short rates moved up from the low levels they attained in the middle of 1971 while long rates remained remarkably steady. This mix of expansionist fiscal policy, moderated slightly by firm (but not tight) monetary policy, is seen as having stimulated the increased real output and fall in unemployment. It is argued however (ch. 2) that, in the absence of the wage-price control program, this stimulation of nominal demand would have had a bigger effect on prices and less effect on real output and employment. Hence, the wage-price controls are seen by the *CEA* as an essential accompaniment to the fiscal and monetary reflation.

The predicted success for 1973 is seen as stemming from a slowing down of the growth rate of nominal GNP from 11 percent to 9 percent resulting from a shift of the budget from a position of deficit to one of balance at full employment, and from a ". . . slower increase of the supply of money. . ." (p. 75). The 9 percent growth in money *GNP* will be translated into a 6 percent (or more) real growth rate and a 3 percent (or less) inflation rate with the help of Phase III of the wage-price controls, the essence of which is ". . . that the Government retains the enforcement ability and authority necessary to the Nation's anti-inflation objective while leaving the private sector the maximum possible freedom to pursue productivity, efficiency, and collective bargaining" (p. 82).

In marked contrast to the successful

* Professor of economics, University of Manchester.
I would like to thank Graham Thomson of the Reserve Bank of Australia and Robert Ward of the University of Manchester for excellent research assistance, and the Social Science Research Council (of the United Kingdom) for financial support. I am especially indebted to my colleagues Malcolm Gray, David Laidler, and George Zis for comments on an earlier draft.

developments of real output, employment, and the price level, the *U.S.* balance of payments remained in deficit in 1972. The balance on current account and long-term capital account was (on a provisional estimate) running at a $10 billion deficit for 1972. This is $1 billion worse than 1971 and $7 billion worse than 1970. The current account alone deteriorated by almost $6 billion during 1972. During this same period the international monetary system has undergone considerable turmoil and change. The Report sees the balance of payments improving during 1973, largely as the result of "... the dollar devaluation, reinforced by the lower rate of inflation in the United States than in other major industrialized countries in 1972, [which] is beginning to affect U.S. exports and imports" (p. 115) but admits that "... long-term changes in trade patterns have tended to make the elimination of the U.S. balance-of-payments deficit more difficult" (p. 116). A crucial ingredient in the solution to America's (and every other country's) balance-of-payments problem is the reform of the world monetary system. The Report presents the, by now, widely rehearsed American case for solving the liquidity problem by placing more emphasis on *SDR*s and for solving the adjustment problem by having greater two-way flexibility of exchange rates based on the objective criterion of reserve changes.

What questions are raised by this brief review of 1972 and preview of 1973 as developed in the Report? The most obvious question which arises concerns the role of the wage-price controls in the slowdown of the inflation rate. The success of American macroeconomic policy in 1972 has not gone unnoticed in other countries. In particular, the British are so enamoured of the American performance that they have set up a wage-price control program which replicates the August 15 *U.S.* measures. Also, discussions underway in Australia as the new Labour Government begins to tackle that country's inflation problem are being heavily influenced by America's recent experience.

The infectiousness of wage-price controls gives some urgency and importance to the question: what was the source of America's success? Did the controls contribute to the slowdown of inflation or was that slowdown entirely due to "... slowing down the rise of Federal spending and continuing the temporary tax increase that had been enacted in 1968 and by tightening monetary conditions" (p. 3). In view of the fact that some countries (again the *UK* in particular) are following the United States with wage-price controls but not with fiscal and monetary stringency, this additional question assumes considerable relevance. Further, there is some danger (hinted at by the Report)[1] that the *U.S.* reflation, which is now underway will, instead of gradually converging on a sustainable degree of resource utilization, overshoot and stimulate a further bout of inflation. If this happens, will a reactivation of measures similar to those used following August 1971 be adequate, or will the economy again have to be put into a 6 percent unemployment recession to get rid of the inflation? These questions are examined in Section I below.

A second set of questions which emerge concern the broad strategy being adopted towards fiscal and monetary policy and the execution of that strategy. Are the demand management policies which are being pursued well conceived and likely to lead to the achievement of "the basic mandate for policy set forth in the Employment Act ... to achieve 'maximum employment, production and purchasing power'?" This is looked at in Section II.

[1] Developments in 1973 have made this danger a reality.

Finally, what are the likely implications, with respect to employment and inflation of the international monetary reforms advocated by the Report? Would those reforms simply free policy making from the balance-of-payments problem, and in themselves, be neutral towards employment and inflation or will the attainable levels of unemployment and inflation be different as a result of introducing *SDR*s as additional reserve assets and adopting greater exchange rate flexibility between the major currencies? This is examined in Section III.

I. The Wage-Price Controls

Did the wage-price controls reduce the inflation rate below what it would have been in the absence of controls? The Report answer is a cautious "yes." It is argued that there are three pieces of evidence which lead to that conclusion. First, the ". . . econometric study of the relation between the rate of inflation and other variables, such as unemployment, prior inflation rates, and growth of demand. [shows that when] The relations derived from past experience are used to project what the inflation rate would have been in the absence of controls. . . . , the projected inflation rate . . . [is] larger than the rate actually experienced in 1972," (p. 61). Second, the distribution by size of wage adjustments in collective bargaining changed importantly between 1971 and 1972. In 1971 the distribution was bimodal with one mode (16 percent of settlements) in the 6–7 percent range and the other (53 percent of settlements) in the open ended 10 percent plus range. In 1972 the distribution became much more compressed and had a single mode (24 percent of settlements) in the 6–7 percent range.[2] Third, evidence from both public attitude surveys and from the behavior of interest rates indicate that inflationary expectations reduced more sharply in the months immediately after August 1971 and continued to decline during 1972 (pp. 62–64).

The first thing that we should note in attempting to evaluate the effects of the controls on the inflation rate is that, although a quarter by quarter explanation of the movements of wages and prices needs a quite sophisticated model, in broader terms, there is no mystery about the determinants of *U.S.* inflation. William Nordhaus has shown that, for annual data, the best simple explanation of the *U.S.* inflation seems to be provided by an expectations augmented Phillips curve.

Nordhaus's equation, estimated on annual data for 1956 to 1968, explains a very high proportion of the variance of the rate of wage change, pp. 443–46, has residuals which appear to be free from autocorrelation, has a coefficient which is not significantly different from unity on the expected inflation rate and predicts the 1969–71 wage change with remarkable accuracy.[3] This simple model has important implications for the cyclical development of inflation and unemployment. A version of the model may be written as:[4]

[2] See the Report, Table 17, p. 62.

[3] Nordhaus's equation is

$$\Delta ln\ w_t = -0.0172 + 1.962\ u_t^{-1} + 0.873\ \Pi_t^e$$
$$\qquad\qquad\qquad (0.371) \qquad (0.204)$$

where $\Delta ln\ w_t$ is the first difference in the natural logarithm of hourly earnings, u_t is the unemployment rate, and Π_t^e is the expected inflation rate (calculated as weighted average of past actual inflation). The figures in parentheses are standard errors. The standard error of estimate is 0.0062 and the *D.W.* 2.36. The prediction errors beyond the end of the sample period were:

1968	1969	1970	1971
+0.8	−0.2	−0.7	+0.6

(see Nordhaus, Table 9, p. 447).

[4] The particular form of the analysis presented here was developed by Herschel Grossman, and suggested to me in discussions with him on this subject. Qualitatively similar formulations are to be found in Edmund Phelps and in David Laidler.

$$(1) \qquad p = \lambda X + p^e$$

$$(2) \qquad p^e = \gamma p + (1 - \gamma) p^e_{-1}$$

where

p = actual rate of inflation
p^e = expected rate of inflation
X = excess demand (proxied by the inverse of unemployment)
$-i$ = time lag

We see that the change in the rate of inflation is given by:

$$(3) \qquad (p - p_{-1}) = \frac{\lambda}{1 - \gamma} X - \lambda X_{-1}$$

In words, the inflation rate rises or falls according to the values of the current and previous levels of excess demand. If excess demand is positive ($X > 0$) and the previous value was negative ($X_{-1} < 0$) then inflation increases. In other words, if the economy is moved from a position of excess supply to excess demand, the inflation rate rises. If, on the other hand, the economy moves from excess demand ($X_{-1} > 0$) to excess supply ($X < 0$), inflation falls. If excess demand is positive in two successive periods, i.e., ($X > 0$, $X_{-1} > 0$), then inflation rises or falls, i.e., $(p - p_{-1}) \gtrless 0$ as $\Delta X / X_{-1} \gtrless -\gamma$. That is, if excess demand falls, unless it falls by more than γ times its previous value, inflation still rises. On the other side, if excess demand is negative, $(p - p_{-1}) \gtrless 0$ as $\Delta X / X_{-1} \lessgtr -\gamma$. That is, with excess supply (negative excess demand), a rise in excess supply always lowers inflation but a fall in excess supply only lowers the inflation rate if it is bigger than γ times its previous value. These qualitative predictions seem to fit the *U.S.* economy almost exactly, provided we assume that $X = 0$ is associated with a 4.5 percent unemployment rate.

Table 1 sets out the annual behavior of the relevant variables in the *U.S.* economy since 1965. Excess demand was positive *and rising* in each of the years 1965, 1966, 1968, and 1969. In each of those years, as predicted, the rate of inflation increased. In 1967, excess demand was steady and the inflation rate was held steady (-0.4). In the years 1970 and 1971, excess demand was negative and increasing, so, and again as predicted, the inflation rate decreased in both years. In 1972, excess demand remained negative but moved closer to zero and the inflation rate steadied. The above analysis is, of course, purely qualitative and I would not want to press it in precise quantitative terms. Nevertheless, in broad terms, it suggests that the *U.S.* inflation—its build-up and gradual decline—can be explained by the movements in the pressure of excess aggregate demand as reflected in the unemployment rate. The question to ask about the wage-price controls then, is not did they reduce the *U.S.* inflation, but did they have any effect in addition to that of the level of aggregate demand?

TABLE 1—THE CHANGES IN INFLATION AND CHANGES IN UNEMPLOYMENT, 1965–72

	Change in Rate of Inflation (1)	Proportionate Change in Unemployment Rate (2)	Sign of Excess Demand	Predicted Sign of Change in Inflation Rate
1965	+0.7	−0.13	0	+
1966	+1.5	−0.16	+	+
1967	−0.4	0.00	+	0
1968	+1.7	−0.05	+	+
1969	+1.4	−0.03	+	+
1970	−0.6	+0.40	−	−
1971	−2.1	+0.20	−	−
1972	0.0	−0.05	−	?

Sources: The Report, Col. (1) Table C-50, p. 262; Col. (2) Table C-26, p. 223.

Let us examine the three pieces of evidence suggested by the Report. First, the econometric work. The only published econometric study seems to be that by Robert Gordon (1972). Gordon first compares his own earlier wage-price model

(1970) with those of Otto Eckstein and Roger Brinner and George L. Perry (1970) and finds his own model tracks the precontrols period better than either of the other two models. He then uses his model to simulate the period from 1971:IV to 1972:II and finds that, on the average during those *three* quarters, wages rose by 0.68 percent per annum and prices by 1.85 percent per annum *less* than the model predicted they should have done.[5] What are we to infer from this? First, of course, we must recognize that we are dealing with a simulated prediction which runs over only three quarters and from a model which, whilst it tracks 1955–71 quite well, is by no means a deterministic relationship. Indeed, the equations were predicting the change in prices in the post-sample precontrols period of 1971:I and II quite badly. The average predicted rate of price change for 1971:I and II was 3.91 percent per annum against an actual rate of 4.65 percent.[6] The prediction error was positive by 1 percent in 1971:I and by ½ percent in 1971:II.[7] Thus, the model was underpredicting price change just prior to the controls but by a decreasing amount. It would therefore be rash to attribute the overprediction during the three controls quarters as being entirely due to the controls.

Second, the fact that Gordon's results show a marked departure of the behavior of prices from that of wages, implying that the controls substantially reduced profit margins should give some doubt about the lasting effects of the controls on the inflation rate. If, over some period of control, two broad indices of inflation do not bear the same relationship to each other as they do on the average, in the absence of controls, then we must expect some compensating departure in the opposite direction, either later in the control period or when the controls are removed. In this case, we must expect some subsequent faster price rises *or* slower wage rises or a combination of the two to restore relative prices to what they would have been in the absence of controls. Whether the adjustment will come out in wages or in prices will depend largely on the monetary and fiscal position. The fact of controls is likely to encourage fiscal and monetary adventurism and therefore to put the economy into a situation such that wages will continue on their course with prices moving ahead more quickly to catch up to their precontrols relative position. Before this issue can be settled, a longer run simulation will be needed taking us through and beyond Phase II and into Phase III.

This evaluation of Gordon's econometric attempt to estimate the effects of the controls on the rate of inflation leads me to accept the view expressed by the Report that ". . . . evidence from this source must be regarded as inconclusive" (p. 61).

The second piece of evidence for the effectiveness of the controls advanced by the Report, which it seems to regard as more convincing than the econometric studies, is the distribution of wage changes in collective bargaining in 1972. I find the reasoning here throughly confusing and misleading. It is true that the distribution by size of wage change shifted downwards considerably between 1971 and 1972. However, the rate of wage and price inflation was also falling at that time, hence, the distribution must have shifted downwards! What is at issue is not the fact but the cause. If the 1972 distribution is compared with 1968 (Table 17, p. 62) it is clear that the differences are minute and entirely in line with the view that the controls had no independent effect on the inflation rate but might have narrowed the

[5] See Gordon (1972), pp. 407–12 and Table 5, p. 411.
[6] See Gordon (1972), p. 408.
[7] These figures are approximate and based on the results charted in Figure 2 of Gordon (1972) p. 409.

TABLE 2—INFLATION AND EXPECTATIONS
(All Percent Per Annum)

	Proxies[a] for Changes in Expected Rate of Inflation over		
	3 months (1)	4–6 months (2)	Long-Term (3)
1971			
August	−0.32	−0.02	−0.05
September	−0.73	0.00	−0.02
October	−0.91	−0.21	−0.25
November	−1.21	−0.83	−0.38
December	−1.38	−1.01	−0.39
1972			
January	−2.00	−1.67	−0.45
February	−2.22	−1.82	−0.37
March	−1.68	−1.58	−0.40
April	−1.68	−1.17	−0.34
May	−1.75	−1.24	−0.34
June	−1.53	−1.11	−0.41
July	−1.34	−0.90	−0.43
August	−1.39	−0.93	−0.45
September	−0.75	−0.61	−0.42
October	−0.68	−0.45	−0.43
November	−0.63	−0.50	−0.52
December	−0.34	−0.30	−0.56

[a] Figures given are differences between the current month's value of the given interest rate and its value in July 1971. Interest rates: Col (1) 3 month U.S. Treasury Bill rate; Col. (2) Prime Commercial Paper 4–6 months rate; and Col. (3) Moody's Corporate Bond Rate (Aaa). Source of basic data: the Report Appendix C, p. 261, Table C-57.

distribution of wage changes by moving people nearer to the guidepost from both directions.

The effects of the controls on expectations are more striking and clear. Table 2 sets out the cumulative movements of three interest rates from August 1971 to December 1972. It is plausible to attribute the sharp fall in short rates by the end of 1971 to a widely held expectation that the controls would reduce inflation by a substantial amount *in the short run*. However, the behavior of long rates indicates that no such expectation was entertained about the longer run. The long rate was falling steadily month by month throughout the entire period of the Phase I "Freeze" and Phase II with no sharp break in its time path in late 1971. Making firm inferences from data like those set out in Table 2 is not possible since many factors, both domestic and international, are at work and affecting the rates. However, the data do not seem to be inconsistent with the view that the controls moved people's short-term expectations about very sharply but had no perceptible impact on long-term inflationary expectations. Since the actual rate of inflation fell by more than 1 percent point over the period from August 1971 to December 1972 a conventional "error learning" hypothesis would be consistent with a fall in expectations as indicated by the long-term bond rate in the region of 0.56 percent points.

Others than the *CEA* and those cited above have done work on the effects of the 1971–72 wage-price controls.[8] Although there is no general consensus in these papers, the points of substance made in them either agree with the position adopted here, or are challenged by the arguments which I have presented. Also these conclusions are entirely in line with those reached concerning the earlier *U.S.* experiments with guideposts and with the experience of other countries.

The effects of the wage guideposts applied in the early 1960's have been extensively studied. George Perry (1967) found some evidence for a downward pressure on wages during that period. However, criticisms of Perry's work by Paul Anderson, Adrian W. Throop and Michael Wachter, and more recently by Gordon (1972) suggest a consensus view (not shared, perhaps, by Perry) that other labor market variables not related to the guideposts, are capable of explaining Perry's results. Philip Cagan et al. (1972)

[8] See, for example, Gardner Ackley (1972), Barry Bosworth, E. R. Fiedler, M. L. Weidenbaum, and H. S. Houthakker (1972a).

also find it impossible to attribute the behavior of wages in the early 1960's to the guideposts.

Evidence from other countries points in the same direction. In particular, evidence from the *UK*,[9] where experience has been more prolonged and varied, indicates a pattern of response to controls which implies temporary success followed by failure such that, on the average, wage and price inflation rates do not depart from the predictions of econometric models, although the quarter by quarter predictions during controls periods are quite erratic as compared with noncontrolled periods. In fact, price equations display more erratic quarter by quarter behavior than the wage equations during such periods. It could well be the case that Gordon's apparently favorable results on prices are influenced by such erratic behavior, which, for the first three quarters of controls happened to reduce the price change. If so, we can expect to see an over-inflation of prices, relative to wage change in subsequent quarters.

In summary then, it seems that the broad movements in the *U.S.* inflation rate can, with little doubt, be attributed to movements in aggregate demand while the independent effects of the controls suggested by the Report (and others) are hard to find, except as transitory phenomena affecting short-run inflationary expectations.

II. Demand Management

The broad strategy of demand management policy[10] has two features. First, the indicator of appropriate policy is presented as the money value of *GNP*; "... the appropriate policy goal in the ensuing period would seem to be a steady increase of money demand (money *GNP*) at a rate consonant with the potential growth rate of the economy and reasonable price stability" (p. 73). It is suggested that a rate of 9 percent is indicated for 1973. How this should be translated into prices, real output, and employment is only suggested and veiled with qualifications. The best guesses are presented as 6 percent (or more) for real output; 3 percent (or less) for prices and unemployment at around $4\frac{1}{2}$ percent. There is clearly, however, no firm commitment, especially to this last figure.

The second feature of the policy strategy concerns the manner of approach to "maximum employment." "... [T]he behavior of the economy in the near future will be determined not only by whether it is operating below or above some numerical potential output or below or above some target rate of employment, but also by the speed and manner with which it approaches the range of its potential levels" (p. 72).

Three questions seem to be raised by this overall strategy for fiscal and monetary policy and by the detailed judgments which lay behind the projections for 1973. First, is the pursuit of some numerical target for the money *GNP* growth rate, rather than some numerical target for the unemployment rate, appropriate? Second, is a gradual approach to full employment a necessary feature of sound demand management policies at the present time? Third, are the predicted movements in the unemployment rate, the inflation rate and level of money *GNP* compatible with each other, and if not, which of them looks like missing its target?

The broad strategy seems to me to be well conceived. It might be described as a sophisticated monetarist approach to demand management. A simple-minded monetarist approach argues that a target growth rate for the money supply will give the maximum stability since any random or other variation in real aggregate spending

[9] See especially Parkin, Sumner, and Jones.

[10] See the Report, ch. 3, pp. 71–74.

will be damped off by associated changes in interest rates. The best objection to this view points out that there are lags in the operation of monetary policy, which lags are presumably present whether such policy is being used actively to initiate a change or passively to damp out changes being initiated elsewhere. On the other hand, fiscal policy operates more quickly and hence variations in aggregate demand might be smoothed out even more completely than they could by monetary policy alone if a money stock expansion rule is augmented by discretionary fiscal policy. However, if fiscal and monetary policy are to be used together and used successfully to stabilize the level of economic activity at or near its maximum output level then, in effect, aggregate demand becomes perfectly inelastic with respect to the general level of prices and wages. In such a situation, the levels of prices and wages and their inflation rates will be determined purely by self-fulfilling expectations on the supply side. This price level inelasticity of aggregate demand can only be avoided if fiscal and monetary policy makers are willing to see some fluctuation in the level of real economic activity and the overall level of employment.

One possible approach is that suggested by the Report, namely to announce a target rate of expansion for money *GNP* and then to leave the private sector, through competitive price fixing and free collective bargaining, to make decisions about wage and price levels in the light of the authorities' announced money *GNP* expansion rate and the underlying trends in real productivity. A tendency to overinflate wages and prices would, in such a situation, result in a rise in the unemployment rate and a slow down in the real expansion rate. A tendency to underinflate wages and prices would have the reverse effect. There would be a built-in mechanism of the expectations augmented Phillips curve type for adjusting the rate of wage and price inflation, albeit at a price. Also, by announcing well ahead of time the planned underlying growth rate of money *GNP and further sticking to those plans*, the authorities can give maximum help to the expectations mechanism. If a target growth rate for money *GNP* is announced and adhered to, and if people come to believe the authorities intentions with regard to money *GNP* expansion, then the implied underlying inflation rate might become widely and firmly expected and hence the underlying actual rate of wage and price adjustment should be broadly in line with that expectation and give rise to minimal unintended adjustments in the overall level of real activity and employment. Controlling money *GNP* has the further virtue that although macro-economic control is less than perfect, control of this particular magnitude is easier than control of either the real component or the price level component separately.

While approving of the broad strategy of monetary and fiscal policy, I have some doubts about the precise details and numerical magnitudes that have been adopted for 1973. A great deal of emphasis is placed by the Report on avoiding "too fast a dash" for the maximum output target on the grounds that the behavior of the economy depends not only on the level of demand but also on the speed with which that level is approached. It is clearly right that a rapid increase in demand, other things equal, will be more inflationary than a gradual increase. It does seem, however, from simulation studies, that provided the economy does not overshoot its natural unemployment rate, then a rapid approach does not have such serious consequences. Recent simulations of the wage price blocks of three large econometric models of the United States indicate that what is most important is not to overshoot. The speed of approach to the full

employment zone whilst not unimportant, is a secondary matter. This is particularly clear in the simulations reported by Albert Hersch.[11] Hersch shows that in a simulation running forward from 1973 through into the 1980's starting out with an inflation rate of 5 percent and an unemployment rate somewhat below 4 percent that if unemployment is allowed to rise for two and one-half years to the neighborhood of 6 percent and then allowed to fall during the next two and one-half years to the neighborhood of $4\frac{1}{4}$ percent the inflation rate will fall off, slowly at first, and then more quickly to the region of $2\frac{1}{2}$ percent. As the unemployment rate settles down at its $4\frac{1}{4}$ percent rate there will, for two years, be a slight increase in the inflation rate, but only by something of the order of $\frac{1}{2}$ percent. The inflation rate will then settle down to a steady level.

Similar implications seem to follow from the monetarist analysis of Leonall Anderson and Keith Carlson. They present simulations in which a monetarist control rule is operated. They use alternative numerical values for the rule, namely 2, 4, and 6 percent for the growth of the money stock. Although the ultimate inflation rates and the maximum unemployment rates *en route* differ between the three time paths, they nevertheless have a similar broad structure. The unemployment rate rises at first, then comes down fairly sharply eventually overshooting slightly and finally approaching a steady inflation rate from above and unemployment rate from below. What seems to be crucial on the basis of these studies is not that the economy approach its full employment target either asymptotically or particularly gradually. An oscillatory approach of the inflation rate seems to be manageable on the basis of past relations. What is crucial is that the economy does not overshoot *and remain in excess demand for any length of time.*

Is there any danger that the present monetary and fiscal policy stance in the United States may push the economy into such a situation? The answer to this question must depend on the way in which the authorities, in detail, conduct their monetary and fiscal policy during 1973. The broad compatibility of the projected inflation and unemployment rates must at least be queried. Most of the large scale econometric models imply that a long-run unemployment rate of $4\frac{1}{2}$ percent is compatible with an inflation rate of somewhat under 3 percent.[12] However, such models tend to ignore the role of inflationary expectations. These are probably running somewhat ahead of the 3 percent level so that for 1973 a somewhat higher unemployment rate would be indicated as being compatible with an inflation rate of the order of 3 percent. The models which imply no long-run tradeoff would also predict a higher unemployment rate being associated with a 3 percent inflation rate, given the inherited expectations of 1973. If the course of the unemployment rate is given most weight then it is likely that money *GNP* will be allowed to grow more quickly than 9 percent and the inflation rate will pickup. If, on the other hand, the money *GNP* growth target is pursued and adhered to, then it seems likely that the unemployment rate will fall less quickly than projected and fail to achieve the $4\frac{1}{2}$ percent target by the end of 1973 while inflation might well be contained at the $2\frac{1}{2}$ percent to $3\frac{1}{2}$ percent range.[13]

III. International Monetary Reform Proposals

The central features of the *U.S.* pro-

[11] See also G. de Menil and J. J. Enzler.

[12] See, e.g., de Menil and Enzler, Table 6, p. 307.

[13] It is now (May 1973) clear that money *GNP* is growing much faster than 9 percent and inflation has taken off again.

posals for international monetary reform are first, the phasing out of gold and phasing in of *SDR*s to operate alongside national currencies (primarily the *U.S.* dollar) as international reserve assets and second, a symmetrical exchange rate adjustment mechanism based on reserve changes. Certain aspects of these proposed changes seem to have received too little attention from the Report.

It has been repeatedly pointed out for two hundred years and most recently by Robert A. Mundell (1971) and Harry G. Johnson (1971, 1973) that, in a world in which exchange rates are rigid, the explanation for inflation has to be sought at the level of world aggregate monetary expansion and excess demand and not at a national (or regional) level. An increase in total demand in the United States does not necessarily lead to an increase in the rate of inflation in that country. If the rest of the world is underemployed, then an increase in American demand would spill over through the balance of payments into an increase in demand elsewhere and have little or no effect on the general inflation rate. If, alternatively, the rest of the world is at full employment then an increase in demand in the United States will lead to the emergence of universal inflationary pressures.

The experience of the world economy in the latter part of the 1960's and the 1970's is an excellent example of such a process. With the *U.S.* balance-of-payments deficit running at $30 billion in 1971 and $12 billion in 1972, and with the rest of the world operating at or near its full employment potential, it is not surprising that the world as a whole has been sharing in the *U.S.* current inflation. The introduction of *SDR*s, at a time when, because of the *U.S.* balance-of-payments deficit, world liquidity was rising too quickly, has simply exacerbated the situation. The initial *SDR* allocations of $3.5 billion in 1970, and $3 billion in each of 1971 and 1972, when added to the *U.S.* balance-of-payments deficits of $10 billion (1970), $30 billion (1971), and $12 billion (1972), have given rise to an increase in world reserves of 18 percent between 1969 and 1970, 45 percent in 1970–71, and 21 percent in 1971–72. Clearly, such increases have inflationary effects.

This is not to argue that *SDR*s are, in principle, a source of inflation. Rather, if world liquidity is to be added to via *SDR* creation, *and if the international community is to exercise control* over the rate of liquidity creation, then the *U.S.* balance-of-payments deficit must be brought much closer to equilibrium than it has been since 1969. To achieve this, *U.S.* monetary policy would have to be modified substantially with a (gradual) slowdown in the rate of money creation possibly combined with exchange rate adjustments. If the United States does not take these steps then it cannot eliminate its inflation. Unlike smaller countries, the United States cannot ignore the effects of its policies on the rest of the world and the feedbacks which arise from those effects. Given the pursuit of full employment policies in all countries outside the United States, a rapid *U.S.* monetary expansion rate and associated balance of payments deficit leads to a world wide inflation in which the United States inevitably shares.

REFERENCES

G. Ackley, "Observations on Phase II and Wage Controls," *Brookings Papers*, Washington 1972, *1*, 173–90.

———, "An Incomes Policy for the 1970's," *Rev. Econ. Statist.*, Aug. 1972, *54*, 218–23.

L. C. Anderson and K. M. Carlson, "An Econometric Analysis of the Relation of Monetary Variables to the Behaviour of Prices and Unemployment," in O. Eckstein, ed., *The Econometrics of Wage Determination Conference*, Washington 1972, 166–83.

P. Anderson, "Wages and the Guideposts:

Comment," *Amer. Econ. Rev.*, June 1969, *59*, 351–54.

B. Bosworth, "Phase II: The U.S. Experiment with an Incomes Policy," *Brookings Papers*, Washington 1972, *2*, 343–83.

P. Cagan, et al., *Economic Policy and Inflation in the Sixties*, Washington 1972.

G. de Menil and J. J. Enzler, "Prices and Wages in the F.R.-M.I.T.-PENN Econometric Model," in O. Eckstein, ed., *The Econometrics of Wage Determination Conference*, Washington 1972, 277–308.

O. Eckstein and R. Brinner, "The Inflation Process in the United States," study prepared for the use of the Joint Economic Committee, 92 Cong. 2d sess., 1972.

E. R. Fiedler, "The Price-Wage Stabilization Program," *Brookings Papers*, Washington 1972, *1*, 199–206.

R. J. Gordon, "The Recent Acceleration of Inflation and its Lessons for the Future," *Brookings Papers*, Washington 1970, *1*, 8–41.

———, "Wage-Price Controls and the Shifting Phillips Curve," *Brookings Papers*, *2*, Washington 1972, 385–421.

A. A. Hersch, "Price Simulations with the O.B.E. Econometric Model" in O. Eckstein, ed., *The Econometrics of Wage Determination Conference*, Washington 1972, 237–76.

H. S. Houthakker, (1972a) "Thoughts on Phase II," *Brookings Papers*, Washington 1972, *1*, 195–98.

———, (1972b) "Are Controls the Answer," *Rev. Econ. Statist.*, Aug. 1972, *54*, 231–34.

H. G. Johnson, "*Inflation and the Monetarist Controversy*," De Vries Lectures 1971, Amsterdam 1972.

———, *Further Essays in Monetary Economics*, London 1973.

D. Laidler, "Simultaneous Fluctuations in Prices and Output: A Business Cycle Approach," *Economica*, Feb. 1973, *40*, 60–72.

R. A. Mundell, *Monetary Theory: Inflation, Interest, and Growth in the World Economy*, Pacific Palisades 1971.

W. D. Nordhaus, "The Worldwide Wage Explosion," *Brookings Papers*, Washington 1972, *2*, 431–63.

J. M. Parkin, M. T. Sumner and R. A. Jones, "A Survey of the Econometric Evidence of the Effects of Incomes Policy on the Rate of Inflation," in J. M. Parkin and M. T. Sumner, eds., *Incomes Policy and Inflation*, Manchester 1972.

G. L. Perry, "Wages and the Guideposts," *Amer. Econ. Rev.*, Sept. 1967, *57*, 397–904.

———, "Wages and the Guideposts: Reply," *Amer. Econ. Rev.*, June 1969, *59*, 365–370.

———, "Changing Labor Markets and Inflation," *Brookings Papers*, Washington 1970, *3*, 411–441.

———, "Controls and Income Shares," *Brookings Papers*, Washington 1972, *1*, 191–94.

E. S. Phelps, *Inflation Policy and Unemployment Theory*, New York 1972.

A. W. Throop, "Wages and the Guideposts: Comment," *Amer. Econ. Rev.*, June 1969, *59*, 358–65.

M. L. Wachter, "Wages and the Guideposts: Comment," *Amer. Econ. Rev.*, June 1969, *59*, 354–358.

M. L. Weidenbaum, "New Initiatives in National Wage and Price Policy," *Rev. Econ. Statist.*, Aug. 1972, *54*, 213–17.

U.S. Council of Economic Advisers, *Economic Report of the President*, Washington, Jan. 1973.

On the Effects of Fiscal and Monetary Policy: A Taxonomic Discussion

By Bent Hansen*

Current debate on monetary and fiscal policies is much concerned with the effects of such policies, and of changes in the budget and money supply. I propose here to discuss some taxonomic problems related to the concept of policy "effects." Their resolution bears directly upon the controversy between Keynesians and Monetarists. I shall show that it is largely a sham-dispute, and it will appear that the empirical findings of the Monetarists have little relation to the Keynesian creed. Even negative effects of the budget with strong positive effects of money supply are fully consistent with strong positive effects of fiscal action and weak or strong effects of monetary action.

The examination of the concept of effects of economic policy will be undertaken in relation to a model that is specified so as to include both the conventional Keynesian set-up for determining effective demand, and a credit mechanism that links effective demand to the banking system in the spirit of the Monetarists without in any sense being non-Keynesian. To remain manageable the model is kept relatively simple and thus cannot do full justice to either side. But it does contain the major traditional fiscal and monetary policy instruments in a closed economy, it permits a crucial institutional specification, and is, indeed, in line with models suggested by sophisticated Monetarists such as Karl Brunner and Allan Meltzer.

The debate has been much concerned with the proper choice of exogenous and endogenous variables in macro-models; and for good reasons. For in economics (as in the experimental sciences), the notion of effect is usually defined as the difference between two states with and without a specified intervention (disturbance), and to intervene in a system means in the simplest case to change the value of an exogenous variable. The effect per unit of discretionary policy measure (or disturbance) is then determined by the multiplier, static or dynamic as the case may be, of the exogenous variable with respect to a specified endogenous variable.[1] To speak about the effects of an endogenous variable, on the other hand, does not make sense.[2]

In the theory of fiscal policy we talk, however, not only about discretionary policy but also about built-in flexibility and make a well-established distinction between effects of discretionary measures

* Professor of economics, University of California, Berkeley. I acknowledge valuable criticism and advice from George Akerlof, George Borts, and, in particular, Carl Christ. There has also been much feedback from work along the lines of this paper, on Indian and Pakistani fiscal and monetary policies pursued by Nurun Nabi Choudhry and Jayanta Roy at Berkeley and from students' comments in class. Werner Neu made the experiments referred to in fn. 16. The paper was read at a seminar at the International Monetary Fund in November 1972.

[1] Jan Tinbergen, p. 2, calls this type of policy "quantitative." In the case of "qualitative" policies the intervention takes on more complicated forms.

[2] This point was first made by Paul Samuelson, pp. 142–43, in his distinction between multipliers and "pseudo-multipliers." It is the same logical problem that mars the use of endogenous variables as "indicators" of monetary policy measures or effects as lucidly demonstrated by Brunner and Meltzer, pp. 196–203.

and effects of built-in flexibility. This distinction, as we shall see, carries over to monetary policy. Conceptually, however, the latter effects differ from the former, as defined in the last paragraph. When talking about effects of built-in flexibility, we have in mind the difference between the states of two different economic structures (viz., economies with budgets that do or do not respond automatically to changes in the economy) exposed to the same disturbance. Since the problem here is to isolate something which can be called the effects of budget flexibility, the two economic structures should differ only in regard to whether changes that occur be classified as discretionary (i.e., exogenous) or automatic (i.e., endogenous). This requirement is easily satisfied. For if a given set of data can be considered as generated by a particular set of relations with a certain set of variables selected as the endogenous variables, that given set of data could, of course, be generated by the same set of relations with another set of variables selected as the endogenous variables. We move thus from one economic structure to the other simply by changing variables. What is automatic response in one interpretation is discretionary action in the other one, and vice versa. Clearly, multipliers defined with respect to the same two variables will differ—maybe even with respect to *sign*, as we shall see—as between two such models, although both models describe the same economy. The "automatic effects" of built-in flexibility per unit of disturbance may then be defined as the difference between two such multipliers defined on the disturbance and the endogenous variable in question.[3]

[3] Richard Musgrave's (1959, pp. 508–10) well-known measure of built-in flexibility is based on the *ratio* between two such multipliers. To make the various effects additive it is preferable to work with the difference rather than the ratio. But that is only a matter of convenience.

This point is crucial for what follows in this paper, and it may be useful to illustrate it by an example related to fiscal policy.[4] Consider the differential form of a simple Keynesian system where Y denotes income, I investment, C consumption, G government expenditure, T tax revenue, and t the tax rate:

$$dY = dI + dC + dG$$

$$dC = c_1(dY - dT) + dc_2$$

$$dT = tdY + Ydt$$

The endogenous variables are assumed to be dY, dC, and dT, while the discretionary policy measures are dG and dt (or, Ydt); as exogenous, noncontrolled disturbances we have dI and dc_2. The solution in dY is

$$dY = \frac{dI + dc_2 + dG - c_1 Ydt}{1 - c_1(1 - t)}$$

and the effects, E_{discr}, of discretionary policy actions are

$$E_{discr} = \frac{dG - c_1 Ydt}{1 - c_1(1 - t)}$$

All sets of empirical observations of dY, dI, dC, dG, dT, dc_2, and Ydt that satisfy the above system will also satisfy the same three relations with dY, dC, and Ydt as endogenous variables, and dG and dT as policy parameters. With the system thus interpreted we have the solution

$$dY = \frac{dI + dc_2 + dG - c_1 dT}{1 - c_1}$$

and the effects, E_{total}, of the total changes in government expenditure and revenue will be

[4] See Hansen (1969, ch. 1, II, A).

$$E_{total} = \frac{dG - c_1 dT}{1 - c_1}$$

With the first interpretation the system has built-in flexibility in tax revenues; but not with the second interpretation. The multipliers and, hence, the effects of both exogenous disturbances, such as dI and dc_2, and of discretionary policy measures, such as dG, differ clearly with the two interpretations. And whereas with the first interpretation it does not make sense to ask for the effects of a change in total tax revenues, dT, this question does make sense with the second interpretation. If we now define the automatic effects, E_{autom}, of built-in flexibility in tax revenues as the difference between the effects of the exogenous, noncontrolled disturbances in the system with and without built-in flexibility in tax revenues, we have

$$\begin{aligned} E_{autom} &= \frac{[dI + dc_2]}{1 - c_1(1 - t)} - \frac{[dI + dc_2]}{1 - c_1} \\ &= \frac{[dG - c_1 dT]}{1 - c_1} - \frac{[dG - c_1 Y dt]}{1 - c_1(1 - t)} \end{aligned}$$

It follows that

$$E_{discr} + E_{autom} = E_{total}$$

a relation that applies to any macrotheoretical set-up and will turn out to be crucial for understanding some of the differences between Keynesians and the Monetarists.

The notion of effects of built-in flexibility calls attention to the fact that the classification of variables as endogenous and exogenous depends entirely upon our "point of view,"[5] that is, upon the questions we seek to answer. We may ask for the effects of disturbances or discretionary measures; we may also ask for the automatic effects of built-in flexibility at disturbances in the economy; and we may ask for the total of discretionary and automatic effects in a given situation. All questions are legitimate in the sense that they can be answered, but they do lead to different classifications of variables. And since they lead to different variable classifications, they imply different multipliers and, hence, effects, too. This simple point seems to have been overlooked by Keynesians as well as Monetarists in the debate of monetary policy.

Another circumstance of some consequence for the choice of variables and of crucial importance for our problem is the so-called budget constraint of the public sector. It was recently brought into the debate by Carl Christ in two important contributions (1967, 1968), but has in fact a long history in the theory of economic policy;[6] in practical monetary analysis it is old hat,[7] and, as we shall see, it has indeed been observed in more careful theoretical contributions (for instance, Brunner and Meltzer). To omit the public sector's budget constraint is a fallacy. Crude Keynesian and quantity theory generally ignores it, however; a typical example is

[5] See Karl Popper, p. 106.

[6] The constraint and its importance was probably first discovered by Knut Wicksell (1898, pp. 6f.) in his macroeconomic analysis of the incidence of taxation and it is generally recognized by modern incidence theory. In monetary theory it has been used by Bertil Ohlin (1934, in particular ch. 5; and 1941, two never translated monetary classics). I myself borrowed the notion from Wicksell and Ohlin to discuss the relation between fiscal policy and money supply (1955), (1958), (1967, chs. 3 and 15). In the early Keynesian literature, Abba Lerner's discussion of "functional finance" (1943), (1946, ch. 24, in particular pp. 314–15) and Lloyd Metzler's wealth study (1951) are the outstanding examples—and exceptions. Some examples from balance of payments theory could also be given. In growth theory Alain Enthoven (1960, p. 315) should be mentioned.

[7] The "bank reserve equation" of the Federal Reserve System is perhaps the best known example; see *The Federal Reserve System, Purposes and Functions*, chs. 12 and 13. A variation on the same theme is the traditional IMF analysis of the "counterpart of money supply changes" that appears in any IMF mission report; see, for instance, *International Financial Statistics*, "Monetary Survey," any country table, any year.

Milton Friedman's recent disclosure (1970) of his monetary system.

I. The Public Sector Budget Constraint

This constraint expresses the fact that net flow of central bank money from the public sector equals net increase of the stock of central bank money held by the private sector and is, indeed, nothing but the change in the central bank's balance sheet between specified points of time. Including creation of central bank money, the public sector is no less constrained in its planning of expenditure and revenue than the household is by *its* budget constraint. Exactly how the constraint should be written depends, however, upon the institutional setting. It would seem that two *Idealtypen* appear in the literature, each one idealizing actually existing arrangements. One of them could be called the European type of constraint, the other one, the United States type of constraint; but with the United States in mind, they could perhaps also be called the pre-Accord and the post-Accord constraints, respectively.

Disregarding foreign transactions (which are easily included), the balance sheet of a central bank that issues currency, keeps deposits (reserves) for and gives loans to commercial banks, and serves as the bank of the government (the Treasury) is

$$L^b + L^g = M^n + M^b$$

where M^n is currency in circulation, M^b commercial banks' deposits at the central bank, L^b rediscounts and advances to commercial banks, and L^g net claims on the government held by the central bank. The change in the balance sheet between two points of time can be written as

$$\begin{aligned}\Delta M^n + \Delta M^b &= \Delta L^b + \Delta L^g \\ &= \Delta L^b + \Delta G - \Delta T + \Delta B \\ &\quad - \Delta B/r_g + D_0\end{aligned}$$

where G is government purchases of goods and services, T tax revenues, B number of government bonds held by the private sector (perpetuaries with \$1 interest payment per period, the only type of claims on the government considered here), r_g bond market rate of interest, and D_0 the budget deficit (government net borrowing) at the beginning of the period. Defining as usual the supply of central bank money, the monetary base, as $M_0 = M^n + M^b - L^b$, we have

$$\Delta M_0 = \Delta G - \Delta T + \Delta B + D_0 - \Delta B/r_g$$

The sum of the first four terms on the right side is the total deficit of the budget (government net borrowing); the last term, $\Delta B/r_g$, is the dollar amount of debt management and open market operations in government bonds.

This form of the constraint could be called the European type of public sector budget constraint; it is the form considered by Christ (1967, p. 436), (1968, p. 56). In a closed economy of this type the supply of central bank money can only change through a deficit (surplus) of the cash budget and through debt management and open market operations; we do not need to distinguish between these two kinds of bond operations provided that all central bank profits are paid to the budget and central bank open market operations are limited to government securities. On the other hand, net cash payments cannot be made to or from the budget or through debt management and open market operations without affecting the supply of central bank money. The budget constraint thus rules out the possibility of studying "isolated" changes in the supply of money or (other) isolated policy measures, and it compels us to specify exactly how money supply changes, through the budget, through operations in the market, or exchange operations (if included), to

mention the three major inroads of money in a modern institutional setting.

We may, however, imagine another ideal type of arrangement in which the government does not at all use the central bank as its banker and thus does not draw upon or keep deposits with the central bank, but to the last nickel finances cash deficits and surpluses through financial transactions with the private sector, in our simplified case through debt management market operations in government bonds. The central bank may, nonetheless, buy and sell government bonds at its own discretion in the open market. We would then have two public sector constraints, one for the central bank which we shall write so that the left side is equal to the change in net central bank money supply, ΔM_0,

$$\Delta M_0 = \Delta M^n + \Delta M^b - \Delta L^b = - \Delta B_{CB}/r_g$$

where $-\Delta B_{CB}$ is the number of government bonds bought by the central bank in the open market, and another constraint for the government (Treasury)

$$\Delta B_{Tr}/r_g = \Delta G - \Delta T + \Delta B + D_0$$

where ΔB_{Tr} is the number of bonds sold by the Treasury to the private sector, and $\Delta B = \Delta B_{Tr} + \Delta B_{CB}$. With these constraints the supply of central bank money can only change through a central bank open market operation; and any budgetary transaction is matched by a corresponding debt management operation in the open market. These two budget constraints represent together what we could call the *U.S.* type; it is the form used by Brunner and Meltzer[8] (1967, p. 208, equation (A 5)). The Accord between the U.S. Treasury and the Federal Reserve System in 1951 could perhaps be described "as if" a shift from the European to the *U.S.* type of constraint took place.[9]

It is immediately clear that what can be done under one of these budget constraints can also be done under the other one. Budget deficit financing through the central bank, which on the European constraint simply means that the treasury draws upon the central bank, must on the *U.S.* model take place through Treasury sales of bonds in the open market to cover the deficit, accompanied by simultaneous central bank purchases in the market to the same amount. The outcome of such concerted discretionary actions will be the same under the two systems. In this sense we might just as well consolidate the two constraints in the *U.S.* type system to the single European type constraint. This, however, would hide the fact that *automatic* responses are different with the two systems. An automatic increase of tax revenue (during an upswing, say) implies with the European constraint an increase of government deposits at the central bank and a corresponding decline of the net supply of central bank money. With the *U.S.* constraint it would leave central bank money supply unchanged but diminish the supply (stock) of government bonds held by the private sector. On the European model the events would imply a "tightening" of the credit market with an upwards pressure on interest rates and a fall in total private wealth; on the *U.S.* model the bond interest rate would fall, while private sector net wealth might fall or increase, and it is more difficult to say what would be the net result. Thus it matters whether a country applies the European and the *U.S.* type of constraint; the effects of built-in flexibility will differ.

[8] In an unpublished paper Brunner and Meltzer (1971) have presented a hybrid form of constraint. See also Brunner (1961).

[9] Strictly speaking the Accord meant, of course, that the Fed stopped automatically supporting the government bond rate at a low level.

In neither Europe nor the United States do we find these *Idealtypen* in pure breed. Some European countries (Germany and Belgium, for instance) are really on the *U.S.* model. And governments always have some working balances with the central bank through which deficits or surpluses may temporarily be financed. In the United States, the Treasury can borrow from the Federal Reserve System to a limit of $5 billion; the Treasury does, on the other hand, keep deposits with the Federal Reserve Banks, although its debt management operations (from the more infrequent bond issues to the weekly Treasury bills auctions) nowadays aim at preventing substantial balances with the Federal Reserve Banks from being built up; in addition, the Treasury simply moves deposits from the Reserve Banks to commercial banks, and vice versa. "By carrying the bulk of its deposits in tax and loan accounts [with commercial banks] the Treasury moderates the effect on bank reserves of fluctuations in its receipts and payments" (*The Federal Reserve System*, p. 194). The intentions are undoubtedly, since the time of the Accord, that it should be up to the discretion of the Federal Reserve Board to decide whether or not it will change its net claims on the government and, hence, the supply of central bank money. These being the intentions it seems justified to approximate the *U.S.* system by assuming the existence of two independent public budget constraints.

In the very short term and for small changes the European and the *U.S.* set-up may not be different; the immediate adjustments may take place on Treasury accounts with the central bank or the commercial banks in either system. But the short term for which this is true may be so short (a few days, a week at most) that it is of little interest in a macroanalysis that centers upon the interaction of real and financial sectors and markets. In the following we shall consider both types of constraints not only because countries are different in this regard, but also, and in particular, because Keynesians generally have tended to think with the European type of constraint in the back of their minds, while American Monetarists naturally have the *U.S.* type of constraint in front of their minds.

II. Model with Fiscal and Monetary Systems

Consider now the following economic relations for an economy with a budgetary and monetary system. The precise specification is not too important for the purpose of taxonomy. We have

$$(1)\qquad Y = I + C + G$$

$$(2)\qquad C = c_1(Y - T + B) + c_2(M^n + \alpha M^b + B/r_g) + c_3$$

$$(3)\qquad I = i_1 r_b + i_2 Y + i_3$$

$$(4)\qquad T = t_1 Y + t_2$$

$$(5)\qquad M^n = nY$$

$$(6)\qquad \alpha M^b = m_1(r_g - \mu r_b) + m_2 Y + m_3 B/r_g$$

$$(7)\qquad \overline{K} + I = (\alpha - 1)M^b + L^b$$

$$(8)\qquad L^b = \beta_1(r_b - \bar{r}) + \beta_2$$

New variables are: $\overline{K}$ is initial private stock of real capital, r_b commercial banks' loan rate, μ is ratio between commercial banks' deposit and loan rates, $\bar{r}$ is central bank's discount rate, α is inverse of required reserve ratio.

We shall assume the signs of constants and parameters to be: $c_1>0$, $c_2>0$, $i_1<0$, $i_2>0$, $t_1>0$, $n>0$, $m_1<0$, $\mu>0$, $m_2>0$, $m_3>0$, $\beta_1>0$ and $B>0$. The choice of signs is in all cases obvious except perhaps $B>0$, there might exist countries with net government claims on the private sector.

Our relations assume that households receive all income and that all private saving is made by households. They own notes, government bonds, and bank deposits but have no bank loans. Enterprises in return finance exclusively through bank loans; this assumption obviously helps to give the banks a crucial role. Total financial wealth of households is thus equal to central bank notes plus bank deposits plus value of government bonds; wealth appears in the consumption function, equation (2), and the demand function for deposits, equation (6), and in the omitted demand function for bonds.[10] We ignore value of ownership, capitalized value of future tax payments, direct benefits of government services, and the like. Enterprises and commercial banks have zero net worth. The investment function, equation (3), contains the commercial banks' loan rate and national income as arguments. Both consumption and investment function are thus highly conventional and acceptable to Keynesians.

The treatment of the banking system suffices to make it an important part of the economy and introduces two important monetary policy instruments, the central bank discount rate and the reserve ratio. Central bank money is divided between currency held by the households and the commercial banks' claims on the central bank, the reserves. The reserves may be counted gross or net because the commercial banks may borrow from the central bank. Demand from households for currency is proportional to national income, equation (5). Equation (6) has on its left side total supply of bank deposits and on its right side total demand for deposits from households. Demand for deposits depends upon the difference between bond rate and deposit rate (the latter assumed to be kept by the banks in fixed proportion, μ, to the loan rate), national income, and households' government bond holdings. We do not distinguish between demand and time deposits and this may be a dangerous simplification (Lyle Gramley and Samuel Chase; Karl Brunner, 1969). The left side of equation (6) may also be read as the inverse of the required reserve ratio times the required, legal reserve, M^b; on this reading the equation determines the legal reserves. The supply of bank loans to enterprises is $(\alpha-1)M^b+L^b$, the right side of equation (7). The term L^b is commercial bank loans (advances or rediscounts) from the central bank or, if negative, the excess reserves of the commercial banks. On the left side of (7) we have the demand for loans from enterprises; it is equal to the (predetermined) total stock of real capital at the beginning of the period plus the investments of the period. We disregard commercial bank investments in government securities (bonds or Treasury bills), another important omission because it prevents us from studying debt management policy in detail. The last equation,

[10] By virtue of Walras' Law we have left out the households' demand function for bonds and the equilibrium condition for the bond market. If we let B^d/r_g denote demand for bonds (in value terms) the equilibrium condition is $B^d/r_g=B/r_g$ and the demand function for bonds is $B^d/r_g=Y[1-(1-t_1)c_1-t_1-n-m_2]-m_1(r_g-ur_b)-t_2(1-c_1)+(1-c_2)(M^n+\alpha M^b+B/r_g)+(1-c_1)B+m_3B/r_g$, which makes sense, given the signs of the constants. In the derivation of this (short-term) demand function for bonds we do not assume that total net worth is given and constant. The instantaneous rate of change of net worth is equal to savings which in our setting may be positive or negative.

The bond market demand equation implied by our model does not suffer from a basic flaw in standard Keynesian models. It is easy to show that, for instance, the standard *IS-ML* model, or Klein's Keynes model assume a kind of Say's Law for bonds so that an increase in the supply of bonds automatically (that is at given income, prices and interest) creates a corresponding increase in the demand for bonds (Hansen (1970, pp. 134–37), Akerlof). Our model thus allows for the "crowding-out" from the financial markets of private borrowers when the government finances a deficit by sales of securities (see Spencer and Yohe where the main references are given).

finally, specifies the commercial banks' demand for central bank loans, or excess reserves as a function of the difference between their own loan rate and the discount rate of the central bank. The higher their own loan rate is, the stronger is the incentive to borrow from the central bank or the weaker is the incentive to keep excess reserves; the higher the discount rate is, the more expensive are central bank loans and the stronger is the incentive to keep excess reserves rather than running the risk of having to borrow from the central bank. These assumptions are in line with current theory.

We have already discussed possible public sector budget constraints and definitions of money supply. All ingredients of both central bank money supply, $M_0=M^n+M^b-L^b$, and money supply in the wider sense, $M_2=M^n+\alpha M^b$, are included in the model and we could add these two equations to the system.

Systems like (1) to (8), completed with a public sector budget constraint, can be used for studying long-term equilibria where the system has come to rest with a balanced budget and a constant amount of money (Christ (1968), also Lerner). It would be entirely unrealistic, however, to assume that the authorities should remain passive during a process toward a long-term stationary state (that may not exist). It is more likely that they would continuously work on rectifying the short-term position. Assuming that the period coincides with the budget year, we shall therefore concentrate on the development from one short-term position to the next and study the effects of economic policy with this short-term perspective, taking for granted that during each period new policy measures will be taken (Christ (1967)). The specification of the available policy instruments may thus conveniently be made in relation to a corresponding differential model (Christ (1967), Hansen (1969, ch. 1, II, A)), determining a change in the short-term equilibrium position. We have

$$(1') \qquad dY = dI + dC + dG$$

$$(2') \qquad dC = c_1dY - c_1dT + c_1dB + c_2dM^n + c_2\alpha dM^b + c_2M^bd\alpha + c_2dB/r_g - c_2\frac{B}{r_g^2}dr_g + dc_3$$

$$(3') \qquad dI = i_1dr_b + i_2dY + di_3$$

$$(4') \qquad dT = t_1dY + Ydt_1$$

$$(5') \qquad dM^n = ndY$$

$$(6') \qquad \alpha dM^b + M^bd\alpha = m_1dr_g - m_1\mu dr_b + m_2dY + m_3\frac{dB}{r_g} - m_3\frac{B}{r_g^2}dr_g$$

$$(7') \qquad dI = (\alpha - 1)dM^b + M^bd\alpha + dL^b$$

$$(8') \qquad dL^b = \beta_1dr_b - \beta_1d\bar{r}$$

We have then to add a public sector budget constraint and, as already pointed out, we have a choice between (at least) two types, describing different behavioral patterns of budgetary and monetary authorities. With the European type of constraint we have

$$(9') \qquad dM^n + dM^b = dG - dT + dB + D_0 - dB/r_g + dL^b$$

whereas the $U.S.$ type of constraint would require the following specification:

$$(9'a) \qquad dM^n + dM^b = dL^b - dB_{CB}/r_g$$

$$(9'b) \qquad dB_{Tr}/r_g = dG - dT + dB + D_0$$

$$(9'c) \qquad dB = dB_{CB} + dB_{Tr}$$

III. Policy Instruments and Endogenous Variables

Thus including a public sector budget constraint of the European type, our system consists of nine equations and we have to select the same number of endogenous variables. Assuming from the outset that in a capitalist economy without direct controls dY, dC, and dI are endogenous variables to be explained by the model, we are left with eleven "policy variables"—dG, dT, dt_1, dB, dr_g, dr_b, $d\bar{r}$, $d\alpha$, dL^b, dM^n, and dM^b—as possible candidates for the other six endogenous variables. The remaining five policy variables are the policy parameters to which effects could be ascribed. From our relations we could thus derive no less than $11!/6!(11-6)!=462$ different models. Even assuming that effects of discretionary measures should only be evaluated in terms of income, dY, there would still be $462\text{x}5=2310$ different effects (multipliers) to calculate; there would, for instance, be 210 different expressions for $dY/d\bar{r}$ (assuming, of course, that all models have sensible solutions). And yet, we have not included the two definitions of money supply which would give us thirteen policy variables from which to choose five.

With a *U.S.* type of constraint we would have eleven equations, and, granted again that dY, dC, and dI be endogenous, we would have thirteen policy variables, the above mentioned plus dB_{CB} and dB_{Tr}, among which to choose eight other endogenous variables. Including the two definitions of money supply we should have fifteen variables to choose among. Our choice is even wider than with the European specification.

If we had no principles for selecting endogenous and exogenous variables, we would have to leave everything like that and be satisfied by presenting all these alternative interpretations and effects. This seems in fact to be what Christ (1967, 1968) recommends. With respect to appraisal of past and present policies, we would thus end up in pure agnosticism; for if we cannot specify what were the policy measures taken, we cannot tell what were the effects of the policies pursued. And with respect to future policy we may give advice that nobody knows how to follow (such as recommending 5 percent increase per year of money supply in the wide sense without specifying the instruments to be used for this purpose).

It should first be noticed, now, that the

Table 1—Summary of Models Investigated

Section Reference	Effects to be studied	Institutional specification, public budget constraint(s)	Equations[a]	Exogenous policy variables	Endogenous variables[b]
IV.A	discretionary	European	(9′)	$G\quad t_1 B \alpha \bar{r}$	T
IV.B	discretionary	*U.S.*	(9′a), (9′b), (9′c)	$G\quad t_1\quad \alpha \bar{r} M_0$	$T\quad B B_{Tr}$
V.A.	total	European	(9), (11′)	$G T\quad B \alpha\quad M_2$	$t_1 \bar{r}$
V.B.	total	European	(9′), (11′a)	$G T\quad \alpha \bar{r} M_0$	$t_1\quad B$
V.C.	total	*U.S.*	(9′a), (9′b), (9′c)	$G T\quad \alpha \bar{r} M_0$	$t_1\quad B B_{Tr}$
VI. {IV.A	automatic	European	{(9′)	$G\quad t_1 B \alpha \bar{r}$	T
{V.A			{(9′), (11′)	$G T\quad B \alpha\quad M_2$	$t_1 \bar{r}$

[a] In addition to equations (1′)–(8′).
[b] In addition to the basic eight endogenous variables, dY, dC, dI, dr_b, dr_g, dM^n, dM^b, dL^b.

introduction of public budget constraints itself limits our choice of exogenous variables somewhat. At least one variable in a constraint must be endogenous, and when only one of the variables in a constraint is endogenous, unacceptable implications may follow. The public budget constraint(s) limit our choice of exogenous variables, but they do not really serve as positive guides for this choice. To this end some "point of view" has to be specified, i.e., some question has to be asked, and some sort of a priori information has to be brought into the picture.

We may, on the one hand, be interested in appraising actual policy actions of the past and present and/or predicting future effects of policy actions to be taken. If so, we have to direct our attention to institutional matters and ask how economic policy actually was, is, or is to be conducted. We have to study past and present procedures and actions of authorities to find out "which variables are under the command of the government" to use Tinbergen's expression (p. 7). With the notable exception of Milton Friedman and David Meiselman (see later) this seems also to be the way most of the participants in the debate look at matters (see, for instance, James Tobin, p. 165).[11]

If, on the other hand, we were interested in problems of budget flexibility and monetary flexibility rather than appraising discretionary policy actions, we would have to consider not only the actual systems with their response patterns but, in addition, an alternative hypothetical system of comparison with different response patterns. Studies of actual administrative systems may help us to choose the hypothetical, comparative system in which somehow budget and monetary flexibility has been removed; but ultimately this choice is a matter of taste.

In this paper we study the effects of discretionary measures in two different models, with European and *U.S.* budget constraints, respectively, and effects of total changes in budget and money supply in three different models, with European budget constraint and money supply defined as central bank money and money supply in the wide sense, respectively, and with the *U.S.* constraint and money supply in the wide sense. Finally we shall show how the automatic effects of built-in flexibility in budgetary and monetary systems can be separated and related to effects of discretionary measures, fiscal and monetary. Table 1 summarizes the characteristics of the models. The discussion is concentrated on the model with European budget constraint in Sections IV.A and V.A. with and without built-in flexibility in budget and money supply.

IV. The Effects of Discretionary Fiscal and Monetary Measures

A. *Model with European Type of Constraint*

We shall now assume that for the particular country and period whose policy actions we want to appraise, the European type of constraint is relevant and that the controlled policy instruments actually were (are) G, t_1, B, $\bar{r}$, and α so that the discretionary measures taken are dG, Ydt_1 (in \$ value), dB, $d\bar{r}$, and $d\alpha$. The endogenous variables are the differentials of Y, C, I, T, r_g, r_b, M^n, M^b, and L^b; see Table 1.

Solving the system for dY, we obtain

$$(10') \quad dY = \frac{1}{\Delta} \text{ [a multiplicand]}$$

[11] Michael J. Hamburger has suggested purely statistical (econometric) criteria for the classification of variables. It is clear, however, that we are confronted here with an analytical problem that cannot be solved by inductive methods; see, for instance, Peter Frost.

The individual terms of the multiplicand are identified in the following tabulation:

I: impact of autonomous change in demand

$dc_3 + di_3 \cdot (\delta_1)$

II: impact of discretionary changes in expenditure and revenue

$+ dG - c_1 Y dt_1 + c_1 dB$

III: impact of discretionary changes in expenditure and revenue via central bank money supply

$+ (dG - Ydt_1 + dB) \cdot (\delta_3)$

IV: impact of initially existing deficit via money supply

$+ D_0 \cdot (\delta_3)$

V: total impact of debt management operations

(i) impact via stock of bonds

$$+ c_2 \frac{dB}{r_g}$$

(ii) impact via bond rate of interest

$$+ \frac{dB}{r_g} \cdot (\delta_2)$$

(iii) impact via central bank money supply

$$- \frac{dB}{r_g} \cdot (\delta_3)$$

VI: impact of change in reserve requirement

$+ M^b d\alpha \cdot (\delta_4)$

VII: impact of change in discount rate

$+ \alpha d\bar{r} \cdot (\delta_5)$

Equation 10′ is a usual multiplier expression; $1/\Delta$ is the multiplier. Our assumptions about the signs of constants and parameters do not suffice for determining the sign of Δ. We assume that it is positive. Apart from autonomous consumption, dc_3, all the exogenous variables appearing in the multiplicand are weighted in various ways. Like Δ, the weights δ_1 to δ_5 are complicated, tedious combinations of the constants and parameters of the model. For our purpose we do not need to write them explicitly. With our assumptions the sign of δ_1 is indeterminate, δ_2 and δ_5 are negative, δ_3 and δ_4 are positive.

We shall comment upon the individual terms of the multiplicand. The expressions inside the multiplicand we call the "impacts"; when multiplying impacts by multiplier we obtain the "effects."

Term I is the impact of autonomous changes of demand. These are the exogenous, nonpolicy disturbances. The impacts of autonomous consumption and investment differ in size, the reason being that in our model investment implies a demand for bank loans, savings a demand for government bonds; investment and savings work thus in different ways on the credit markets. Note that we cannot be sure that the impact of autonomous investment is positive.

Term II is the direct impact on demand of discretionary fiscal measures dG, Ydt, and dB. The term dB is the increase of government interest payments related to the period's bond sales. It seems natural to include changes in interest payments with other fiscal measures, although, strictly speaking, they are the outcome of debt management and open market operation decisions. Terms I and II will be recognized as good "Keynesian" terms (with a small reservation for autonomous investments)[12] working directly on effective demand.

[12] The weight, δ_1, of autonomous investment has the form $1+\epsilon_1$ where ϵ_1 (negative) is a combination of the constants α, β_1, i_1, m_1, m_3, n, c_2, and B/r_g^2, all of them re-

Term III is particularly interesting from our point of view. It shows the impact of all discretionary fiscal policy measures via the change in supply of central bank money which they directly imply. Since this impact works through changes in money supply it might seem reasonable to call Term III the monetary impact of discretionary fiscal measures and to count it as part of monetary policy. Keynesians might, nevertheless, want this impact to be credited to fiscal policy; it arises after all only because fiscal measures are taken and is, indeed, a kind of mirror image of these measures. The sign of this term is positive; thus, its inclusion among the effects of fiscal policy would certainly strengthen the Keynesian position.

Term IV is, of course, the Lerner-effect of "functional finance." If there is a cash deficit in the initial position, central bank money supply will increase automatically with this amount. This is a predetermined effect from the budget that will exist as long as the budget is not balanced. We are up against the same classification problem as with Term III. The sign is positive and the term may be important, particularly if the deficit is large. Terms III and IV could be consolidated into $(dG-Ydt_1+dB+D_0)\delta_3$; this term would be the impact from the change in central bank money supply implied by initial deficit and discretionary fiscal measures together. Note, however, that this consolidated term does not include automatic changes in tax revenues and thus cannot be described as the impact of the total deficit. Nor is it equal to the total change in central bank money supply; even if we included $-dB/r_0$ (from term V, iii (see below)), it would still not include those changes in central bank money supply that are induced by automatic tax revenue changes.

Term V collects the impact of debt management and open market operations; analytically there is no difference between these two kinds of operations. The term is split up into three subterms according to the way through which the impact is exerted. There are impacts via the change in the stock of bonds (i), via the accompanying change in households' cash balances (iii),[13] and via the capital gains or losses associated with changes in the bond rate (ii). While the sign of Term V as a whole is indeterminate, the sign of the subterm (i) is positive, and those of subterms (ii) and (iii) are negative. Thus we cannot say a priori whether open market operations are expansive or contractive. How should this term be allocated in relation to fiscal and monetary policy? To exclude the open market operations of the central bank from monetary policy would seem odd; to split the term in two parts according to whether the operations are carried out by central bank or Treasury does not make much sense, either, because the effects are identically the same.

Terms VI and VII are certainly the outcome of monetary policy. They are the impact of reserve requirement and discount rate changes. The signs of the weights of these monetary measures are positive and negative, respectively, as should be expected.

Any attempt to classify the effects of policy measures thus runs into ambiguities. At one extreme we have a classification which counts only Term II as fiscal policy

lated to the monetary system. The impact of autonomous investment could, thus, be broken down in $di_3+di_3\epsilon_1$ of which the first term is the direct "Keynesian" impact on effective demand, and the second one is the impact through monetary repercussions and might be counted as part of the monetary effects.

[13] Lerner, p. 311, pointed out that the effects of taxes and borrowing cannot be clearly separated because, in addition to taking income and adding to people's assets, respectively, both imply a drain on money supply and thus overlap to some extent. This is precisely the reason why we find Ydt_1 in both Terms II and III, and dB/r in both Terms V (i) and V (iii); Ydt_1 in Term III and $-dB/r_0$ in V (iii) have the same weight, δ_3.

impact, and considers all the other terms from III to VII as impacts of monetary policy;[14] and at the other extreme, one that classifies Terms II, III, IV, and V as fiscal (or, at least, budgetary) policy impacts, and only leaves Terms VI and VII as monetary policy impacts. And in between we have several alternatives to choose amongst. Even if we single out Term V and call it the impact of debt management and open market operations, and include the latter in monetary policy, we are still left with the problem of classifying Terms III and IV. The most natural reaction to these ambiguities might be simply to give up the idea of classifying everything into fiscal and monetary policy, and work with all the individual terms of the multiplicand that model and selection of policy instruments give rise to.

Does money matter in this model? This question may be interpreted in two different ways. Does it matter whether there exists a monetary system, and how it is organized and how it functions? Does it matter, given the monetary system, whether the quantity of money is larger or smaller, or changes? On both interpretations money certainly matters in this model, although we cannot ascribe definite, well-defined effects to the total changes in money supply; but we have well-defined "liquidity effects."

B. *Model with U.S. Type of Constraint*

We use here equations (1′) to (8′) and (9′a), (9′b), and (9′c). We shall assume that the discretionary measures are dG, dt_1, $d\alpha$, $d\bar{r}$, and $dB_{CB}/r_g(=-dM_0)$. For the endogenous variables, see Table 1. In terms of policy measures the difference between the two types of models is that in the European type model all transactions in government bonds, dB/r_g, were considered a discretionary measure; in the present model only the central bank open market operations in bonds have this position. Treasury operations, dB_{Tr}/r_g, are endogenous and do not have well-defined effects. The implication is that changes in the supply of central bank money can be considered a discretionary policy measure with well-defined effects equal to, and indistinguishable from, the effects of open market operations (with opposite sign).

Solving for dY we now get

$$(10'\text{a}) \quad dY = \frac{1}{\Delta'}\,[\text{a multiplicand}]$$

the terms of the multiplicand being:

I: direct impact of autonomous changes in demand

$$dc_3 + di_3(\delta_1')$$

II: direct impact of discretionary budgetary measures

$$+\,dG - c_1 Y dt_1 + c_1 dB_{CB}$$

III: impact of discretionary budget changes and initial budget deficit via implied increase of stock of bonds

$$+\,(dG - Ydt_1 + dB_{CB} + D_0)(\delta_3')$$

IV: impact of change in reserve requirements

$$+\,M^b d\alpha(\delta_4')$$

V: impact of change in discount rate

$$+\,\beta_1 d\bar{r}(-\alpha\delta_4')$$

VI: impact of open market operations

$$+\,dB_{CB}/r_g(\delta_6')$$

The multiplier is $1/\Delta'$, assumed to be

[14] Even part of the impact of autonomous investment could be included here; see fn. 12.

positive. The terms δ' are weights (different from the δ in (10′)); with our assumptions δ'_3 and δ'_4 are positive and δ'_5 is negative, while the sign of δ'_1 is indeterminate.

We note first that the impacts of both government spending (dG) and increased central bank money supply ($-dB_{CB}/r_0$) are positive: a priori we cannot say which impact is the largest one.

The division between effects of fiscal and monetary policies is simpler in this case than with the European type constraint. Not even Term III really gives rise to ambiguity. This (positive) term shows the impact of the increase of private sector bond holdings, accompanying an increase of expenditure or a fall in revenue. Note that this term cannot be considered the impact of debt management operations because it does not show the impact of *sales* of bonds against money (Term VI shows the impact of such sales), and Keynesians might rightly claim credit to fiscal policy for this term; without a budget deficit or surplus this change in government bonds with the private sector would not take place. Since Term III is definitely positive its inclusion in fiscal policy will again strengthen the Keynesian position.

V. The Total Effects of Fiscal and Monetary Policy

We turn now to the other point of view—the problem of built-in flexibility and its consequences. For that purpose we have first to set up a system which is equivalent to the system consisting of equations (1′) to (8′) (with the relevant budget constraint(s)) in the sense that a given set of data which could have been generated by the first system could just as well have been generated by the new system, but differs from the first system in having no built-in flexibility, either in the budget items or in money supply. On the basis of such a system we can consider the *total* changes in government expenditure and revenue, and in the supply of money as government policy measures and study their effects (briefly, the total effects); and we shall use it in Section VI for defining the automatic effects of built-in flexibility in budgetary and monetary system in line with the introductory remarks.

We have clearly to consider T as an instrument and dT as a policy measure. Consequently, we have either to discard equation (4′) or consider dt_1 as an endogenous variable. These two methods of procedure are equivalent. With respect to debt management and monetary policy we shall distinguish between the European and the *U.S.* type of constraints.

A. *Model with European Type of Constraint and M_2 as Policy Variable*

We continue to interpret dB and dB/r_0 as exogenously determined; otherwise, there would be some built-in flexibility even in the current budget via interest payments. A moment's reflection makes it clear that it is not possible to remove all built-in flexibility with respect to the monetary system. For if no automatic changes were to take place in the balance sheet of the central bank, and the reserve requirements as well as the discount rate were still to be considered as controlled by the central bank, we would have no less than eight exogenous variables, namely—in addition to dG, dT and $dB-dL^p$, dM^n, dM^b, $d\bar{r}$, and $d\alpha$. The model would in return contain only six endogenous variables and would be overdetermined. With respect to the remaining three endogenous variables the choice is rather arbitrary, but in line with current debate it would be natural to let either central bank money

or money supply in the wider sense appear as policy parameters. Since Monetarists are vacillating between these definitions of money supply (and others) it may be appropriate to study the implications of shifting from one money supply to the other as exogenous policy variable.

Choosing first money supply in the wider sense as the exogenous variable, we add to our equation system a new equation:

$$(11') \quad dM_2 = dM^n + \alpha dM^b + M^b d\alpha,$$

which we obtain from $M_2=M^n+M^b\alpha$. We consider dM_2 as an exogenous policy measure and dM^n and dM^b as endogenous variables, so that the composition of the change of money supply is endogenously determined. In addition, we shall choose to let commercial banks' borrowing, dL^b, and the discount rate, $d\bar{r}$, appear as endogenous. To let the central bank's discount rate appear as endogenous is, of course, unusual in the sense that the discount rate now appears as a kind of "market" rate; it just means, however, that the central bank "follows" the market. In any case, this one is an "as if" model, where we are only interested in getting built-in flexibility removed from budget and supply of money, and we remove it at the cost of assuming built-in flexibility in marginal tax rate, i_1, and discount rate, $\bar{r}$.

We thus work with the system consisting of equations (1′) to (8′), (9′) and (11′). The policy measures are dG, dT, dB, $d\alpha$, and dM_2. Endogenous variables are shown in Table 1. It is clear that a set of data generated by the present model could also have been generated by model (1′) to (8′) and (9′). This is so because we could always add equation (11′) to that model and interpret dM_2 as an endogenous variable.

Solving for dY we now obtain

$$(10^*) \quad dY = \frac{1}{\Delta^*}\,[\text{a multiplicand}]$$

the terms of the multiplicand being:

I: impact of autonomous changes in demand

$$dc_3 + di_3(\delta^*\mu m_1)$$

II: impact of total changes in expenditure and revenue

$$+\, dG - c_1 dT + c_1 dB$$

III: impact of total cash deficit via change in central bank money supply

$$+\,(dG - dT + dB + D_0)(\delta^*\mu m_1 - 1)$$

IV: impact of debt management and open market operations

$$+\,\frac{dB}{r_g}\,(1 + c_2 - \delta^*(\mu m_1 - m_3 i_1))$$

V: impact of change in money supply in wide sense

$$+\, dM_2(1 + c_2 - \delta^*(\mu m_1 + i_1))$$

where $\Delta^* = 1 - c_1 - \delta^*(i_2\mu m_1 + i_1(n + m_1))$; we assume Δ^* to be positive. In the weights of the multiplicand we have the common factor

$$\delta^* = \frac{c_2}{m_4 i_1}\,\frac{B}{r_g^2}, \text{ sign positive}$$

and

$$m_4 = m_1 - m_3\,\frac{B}{r_g^2}, \text{ sign negative}$$

We first note a peculiarity: changes in reserve requirements, $d\alpha$, do not appear in the multiplicand, although $d\alpha$ belongs to our policy measures, the reason being that it has zero effects. Its effects are, as it were, absorbed in the (endogenous) changes of

the discount rate; and changes in the discount rate have no effects because it is endogenous. Thus the Monetarist would be justified in concentrating on the effects of changes in money supply in the wider sense.

A second striking feature is that autonomous investments, di_3, enter with a *negative* weight (m_1 being negative). This remarkable result is related to the position of the banking system with the assumption of money supply, dM_2, being exogenous. An increase in investments requires an increase in bank loans. Given the reserve ratio, bank loans can increase if M^b increases which at given dM_2 and α can only happen if currency circulation falls, which in turn requires that income fall. Bank loans can also increase at given M_2 if bank borrowing from the central bank, L^b, increases, but this requires an increase in interest rates, which will again help to keep back investments and via wealth accomplish a fall in consumption. This fall in consumption must be so strong that we get a fall in income despite some expansion of investments. It should be emphasized that this result does not hinge upon the magnitude of responses to interest rate changes, but only upon their signs.

Looking at Terms II and V we see that both government spending and money supply have positive weights, but while that of spending, dG, is equal to one, that of money supply>1. Transfers (dT and dB) have the weight $c_1<1$. A priori, we can thus state that the effects of money supply exceed those of changes in budget expenditure and revenue. And things get even worse for the Keynesians when we note that in Term III we have the monetary impact of the budget deficit through changes in central bank money supply implied by the deficit, and the weight of this term is negative and <-1. If we consolidate Terms II and III we get

$$dG(\delta^*\mu m_1) - (dT - dB)(c_1 + \delta^*\mu m_1 - 1) + D_0(\delta^*\mu m_1 - 1)$$

where both dG, $-(dT-dB)$, and D_0 appear with negative weights. What is usually considered expansive fiscal policy thus implies a fall in income. With the monetary effects of budgetary changes and the deficit included in the effects from the budget, we thus end up with negative effects of fiscal policy and positive effects of changes in money supply in the wide sense. Which settles matters definitely in favor of the Monetarists.

Again it may be asked why we obtain the remarkable result that Keynesian "expansive" policies, a budget deficit, increased spending and a fall in taxes all have negative effects on income. And again the key to the answer is in the assumption that money supply in the wide sense is kept constant despite the increase of central bank money supply. A glance at equation (9′) shows immediately that if for instance dG is positive ($dT=dB=D_0=0$) so that central bank money increases via the budget, then at given note circulation, M^n, and bank borrowing, L^b, bank reserves, M^b, must increase by an amount equal to dG through an increase of deposits. The value of M_2 has thus increased by dG and to prevent a further increase through the bank's expanding their loans on the basis of the reserve increase, the central bank rate has to increase so much that commercial bank loans from the central bank fall, and/or excess reserves are built up by the amount dG. But this still leaves the nonbank sector with the initial increase in deposits equal to dG and at given required reserves bank loans to investments (see equation (7′))have to be cut down by this amount. Thus investments will fall by dG and the expansionary effects of the increase in government ex-

penditure through the Keynesian part of the system will be completely offset: there is no increase in income. This is the "crowding out" mechanism that the St. Louis school emphasizes (Spencer and Yohe). However, the only way (in our model) for the banks to adjust loans downwards is through an increase in the bank loan rate which through substitution between deposits (including time deposits) and bonds will pull up the bond rate. There is thus a fall in private wealth: consumption and, hence, income will now fall. "Crowding" in the credit market with wealth effects on consumption are thus ultimately responsible for the negative effect on income of an increase in government expenditure. There is nothing mysterious in this result.

Finally we have in Term IV the impact of debt management and open market operations. It can be broken down in subterms as we did in (10′). The weight has an indeterminate sign. We are left with the problem of whether this term should be counted as monetary or budgetary policy, be kept separate as debt management policy, or be divided according to who makes the operations.

To make the best possible of the situation for the Keynesians we should consider Term III as monetary policy impact which, if needed, could be neutralized by appropriate debt management or open market operations. The impact of changes in expenditure and revenue is then of the order of magnitude usually claimed by Keynesians (albeit definitely smaller than the impact of a change in money supply). But it should then be noted that irrespective of whether fiscal policy is expansive or contractive we have simultaneous monetary effects (partly predetermined) working through the budget, and these effects are "perverse," that is, contractive or expansive depending upon whether there is a total deficit or surplus.

B. *European Constraint and M_0 as Policy Variable*

If instead of M_2 we had chosen net central bank money supply,

$$(11'a) \qquad M_0 = M^n + M^b - L^b$$

as the exogenous policy variable, we would be forced to let dB/r_g appear as endogenous because the public sector budget constraint would be $dM_0 = dG - dT + dB - dB/r_g + D_0$. At least one of the variables in the constraint must be endogenous, and with D_0 and r_g predetermined and dM_0, dG, and dT exogenous, we have no other choice than to let dB be endogenous. This means that we have some flexibility in the budget, viz. *via* the interest payments; and we cannot speak about effects of debt management and open market operations. Since interest payments and their flexibility may be considered a secondary problem, we might be willing to accept working with dB as endogenous. Policy variables and endogenous variables are shown in Table 1. The solution for dY is then

$$(10^{**}) \quad dY = \frac{1}{\Delta^{**}}\,[\text{a multiplicand}]$$

the terms of the multiplicand being:

I: impact of autonomous changes in demand

$$dc_3\delta_1^{**} + di_3(1 - \delta_2^{**})$$

II: impact of total changes in expenditure and tax revenue

$$+\, dG - c_1 dT$$

III: impact of total deficit (disregarding interest payments) via stock of bonds

$$+\, (dG - dT + D_0)\delta_3^{**}$$

IV: impact of reserve requirement and discount rate changes

$$+ (M^b d\alpha - \alpha\beta d\bar{r})\overset{**}{\delta_4}$$

V: impact of changes in central bank money supply

$$+ dM_0(\alpha\overset{*}{\delta_4} - \overset{*}{\delta_2} - \overset{*}{\delta_3})$$

where it can be shown that all δ are positive assuming $-m_1 < \alpha\beta/\mu$. The weight of dM_0 is positive and so is the weight of dG. A priori we cannot say which one is largest. Interest payments do not appear in Terms II and III and there is no term, of course, for impact from bond transactions. Once more, we have the problem of how to classify Term III. And we note that now both reserve requirement and discount rate changes appear with monetary policy impacts besides money supply. At comparison between (10*) and (10**) we find that while the effects of money supply in the wide sense definitely exceed those of fiscal policy (which even become negative when monetary effects are included), it is not possible a priori to say whether the effects of central bank money supply exceed or fall short of those of fiscal policy. Monetarists will be strikingly right if we concentrate on money supply in the wide sense; Keynesians may be right if we concentrate on central bank money supply.

C. *Model with U.S. Type of Constraint*

Things work out somewhat differently with the *U.S.* type of budget constraint, as could be expected. Since the institutional setting here is arranged in advance so as to prevent flexibility in central bank money supply, we do not need to ask how to remove this kind of flexibility. The Accord took care of that (by and large). On the other hand, if we insist on keeping the characteristic feature of the budget constraint unchanged, viz. that cash deficits or surpluses always be financed through debt management operations with the private sector, we have to continue to let debt management operations, dB_{Tr}/r_g, be endogenous; thus we cannot remove budget flexibility with respect to interest payments. Quantitatively this may perhaps be considered a secondary matter. We have, then, three possibilities with regard to the set-up of exogenous variables:

(a) We can select dG, dT, $d\alpha$, $d\bar{r}$, and $dB_{CB}/r_g(=-dM_0)$ as exogenously determined policy variables. The only difference to the previous solution (10′a) of the *U.S.* type of model is that flexibility in tax revenue is now removed.

(b) We can select dG, dT, $d\alpha$, dB_{CB}/r_g $(=-dM_0)$, *and* dM_2. In this case both central bank money supply and money supply in the wide sense are exogenously determined policy variables; no Monetarist has, however, had such a possibility in mind (for instance, discussing the effect of an increase in the monetary base at unchanged money supply in the wide sense). We shall leave this possibility aside.

(c) We could, finally, select dG, dT, $d\alpha$, $d\bar{r}$, and dM_2, implying that we remove flexibility in money supply in the wide sense, but permit flexibility in central bank money supply. To reintroduce flexibility in M_0 means, however, switching to the European type of constraint, we have already studied this case (equation (10*)).

We have thus in effect only one case, (a), to consider with the *U.S.* type of constraint. We work accordingly with a model consisting of equations (1′) to (8′), and (9′a), (9′b), and (9′c), and assume that the exogenous policy variables are dG, dT, $d\alpha$, $d\bar{r}$, $dB_{CB}/r_g(=-dM_0)$. The endogenous variables are shown in Table 1. The solution for dY is

$$(10^{*a})\quad dY = \frac{1}{\Delta^{*a}}\,[\text{a multiplicand}]$$

the terms of the multiplicand being:

I: impact of autonomous changes in demand

$$dc_3 + di_3(\delta_1^{*a})$$

II: impact of total changes in budget expenditure and revenue

$$+ dG - c_1 dT + c_1 dB_{CB}$$

III: impact of total cash deficit via implied increase of private sector bond holdings

$$+ (dG - dT + dB_{CB} + D_0)(\delta_3^{*a})$$

IV: impact of changes in reserve requirements

$$+ M^b d\alpha(\delta_4^{*a})$$

V: impact of changes in discount rate

$$+ \beta_1 d\bar{r}(-\alpha\delta_4^{*a})$$

VI: impact of open market operations (change in central bank money supply with opposite sign)

$$+ \frac{dB_{CB}}{r_g}(\delta_5^{*a})$$

with $1/\Delta^{*a}$ being the multiplier (again, too tedious to write down) assumed to be positive. The coefficients δ_3^{*a} and δ_4^{*a} are positive, δ_5^{*a} is negative, while the sign of δ_1^{*a} is indeterminate.

Note that Terms II and III do not include change in interest payments arising from Treasury transactions in bonds, and strictly speaking these terms do not comprise the *total* changes in expenditure and revenue and the deficit. The interpretation of Term III is the same as that of Term III in (10′a). Both Term II and VI are positive, and we cannot say a priori whether a change in budget expenditures and revenue or central bank money supply has the strongest effect, even if we include Term III in fiscal policy.

Note also that our results in this section and the previous ones are at variance with those of William Silber.

VI. The Effects of Built-In Flexibility in Budgetary and Monetary System

By the automatic effects of built-in flexibility of budget and monetary system together we shall, in line with the remarks in the introductory paragraphs, understand the difference between the effects of noncontrolled disturbances on national income at the actual budgetary and monetary systems, and at a hypothetical system with no built-in flexibility.

To fix ideas, we shall assume that system (1′) to (8′) with the European constraint (9′) and dG, dt_1, dB, $d\alpha$, and $d\bar{r}$ as policy parameters accurately describes the actually existing system and its administrative *modus operandi*. System (1′) to (8′) with (9′) and (11′) and dG, dT, dB, $d\alpha$, and dM_2 as exogenous policy variables is then a hypothetical system without built-in flexibility in budget and money supply in the wide sense. (See Table 1.) Instead of tax rate and discount rate we let tax revenue and money supply play the roles as policy parameters.

If we let E_{autom} denote the automatic effects of all built-in flexibility thus specified, we have from solutions (10′) and (10*), dc_3 and di_3 being the nonpolicy disturbances,

$$(12)\quad E_{autom} = \frac{1}{\Delta}\,[dc_3 + di_3\delta_1] - \frac{1}{\Delta^*}\,[dc_3 + di_3\delta^*\mu m_1]$$

From (10′) and (10*) it follows that we could also write

$$(12') \quad E_{autom} = \frac{1}{\Delta^*}\Big[dG - c_1 dT + c_1 dB + \frac{dB}{r_g}(1 + c_2 - \delta^*(\mu m_1 - m_3 i_1)) + (dG - dT + dB + D_0) \cdot (\delta^* \mu m_1 - 1) + dM_2(1 + c_2 - \delta^*(\mu m_1 + i_1))\Big] - \frac{1}{\Delta}\Big[dG - c_1 Y dt_1 + c_1 dB + \frac{dB}{r_g}(c_2 + \delta_2 - \delta_3) + (dG - Y dt_1 + dB + D_0)\delta_3 + M^b d\alpha(\delta_4) + \alpha\beta_1 d\bar{r}(\delta_5)\Big]$$

We want to split E_{autom} into automatic effects of the budget and the monetary system, respectively. Two problems arise. How shall we deal with debt management operations? And once again we are confronted with the question of whom to credit for the monetary implications of fiscal action. To exemplify, we shall here show the definitions that follow from including monetary implications of fiscal actions and budget changes in fiscal policy, but leave debt management and the monetary implications of the initial deficit in monetary policy.

We then naturally define the automatic effects of the budgetary system, $E_{aut.budg.}$, as

$$(13) \quad E_{aut.budg.} = \frac{1}{\Delta^*}[dG\delta^*\mu m_1 - (dT - dB) \cdot (c_1 + \delta^*\mu m_1 - 1)] - \frac{1}{\Delta}[dG(1 + \delta_3) - (Y dt_1 - dB)(c_1 + \delta_3)]$$

If, moreover, we agree that the automatic effects of the monetary system, $E_{aut.mon.}$, should be defined so that $E_{aut.mon.} + E_{aut.budg.} = E_{autom}$, we have from (12′) and (13)

$$(14) \quad E_{aut.mon.} = \frac{1}{\Delta^*}\Big[\frac{dB}{r_g}(1 + c_2 - \delta^*(\mu m_1 - m_3 i_1)) + D_0\delta^*\mu(m_1 - 1) + dM_2(1 + c_2 - \delta^*(\mu m_1 + i_1))\Big] - \frac{1}{\Delta}\Big[\frac{dB}{r_g}(c_2 + \delta_2 - \delta_3) + D_0\delta_3 + M^b d\alpha(\delta_4) + \alpha\beta_1 d\bar{r}(\delta_5)\Big]$$

Defining, furthermore, the effects of discretionary budgetary measures, $E_{discr.budg.}$, and monetary measures, $E_{discr.mon.}$, as (see 10′)

$$(15) \quad E_{discr.budg.} = \frac{1}{\Delta}[dG(1 + \delta_3) - (Y dt_1 - dB)(c_1 + \delta_3)]$$

and

$$(16) \quad E_{discr.mon.} = \frac{1}{\Delta}\Big[\frac{dB}{r_g}(c_2 + \delta_2 - \delta_3) + D_0\delta_3 + M^b d\alpha(\delta_4) + \alpha\beta_1 d\bar{r}(\delta_5)\Big]$$

and recalling that the effects of nonpolicy disturbances when there is no built-in flexibility in budget or money supply in the wide sense, $E_{non-policy}$, is

$$(17) \quad E_{non-policy} = \frac{1}{\Delta^*}[dc_3 + di_3\delta^*\mu m_1]$$

we have the total change in national in-

come broken down upon the (total) effects of nonpolicy disturbances, discretionary fiscal and monetary policy measures, and automatic effects from built-in flexibility in budgetary and monetary systems, thus

$$(18)\quad dY = E_{non-policy} + E_{discr.budg.} + E_{discr.mon.} + E_{aut.budg.} + E_{aut.mon.}$$

Finally, we remind the reader about the two basic ambiguities in this definitional system: the removal of flexibility in budget and monetary system can be made in several ways, and the individual impacts can be distributed as fiscal and monetary policy differently from what we have done here. An obvious possibility is to single out bond sales as a separate policy (which will then have both discretionary and automatic effects) or, perhaps, distribute them on budgetary and monetary authorities. Similar expressions could be worked out for the model with the United States type of constraint. Since our main concern is concepts and definitions, we shall not enter upon the problem of the size of automatic effects in different systems although this must be an important consideration in the choice of institutional system.

VII. Some Implications for the Friedman-Meiselman Controversy

In their attack on Keynesian theory, Friedman and Meiselman (1963), (later referred to as F-M), tested the following hypotheses (we leave out the price variable which is of no relevance for what follows):

$$(19a)\quad C = a + VM$$

or

$$(19a')\quad dC = da + VdM$$

$$(19b)\quad C = a + KA$$

or

$$(19b')\quad dC = da + KdA$$

$$(19c)\quad C = a + VM + KA$$

or

$$(19c')\quad dC = da + VdM + KdA$$

thus explaining consumption by "money only" ((19a) in which V is velocity), by "autonomous expenditure only" ((19b) in which A is autonomous expenditure and K the multiplier), and finally by "both money and expenditure" (19c). F-M chose to use C as a dependent variable for purely statistical reasons. A number of definitions of M were tried out; they ended up with a definition corresponding to our M_2 (currency in public circulation plus adjusted demand deposits plus time deposits in commercial banks; see F-M, (1963) p. 184). A was defined as private domestic investment *plus* the government deficit on income and product account, *plus* the net foreign balance.

F-M's methods were criticized partly for misspecification, partly for statistical reasons. Here we shall comment mainly on the first problem.

F-M (1963, p. 174) were, of course, fully aware that the hypotheses to be tested depend upon the specifications of the basic structural models of which the second and third equations are reduced forms. There is little point in demonstrating this acknowledged fact. In deriving the reduced forms, however, nobody tried to derive the combined form from a basic structural model, and it must, of course, be the combined form upon which interest should center. Albert Ando and Franco Modigliani did set up a model containing both budget variables and money supply, but the model was mainly used for discussing the conditions for the extreme first and second hypotheses (19a) and (19b) to follow from the model as reduced forms. Thus, it may perhaps be of some interest

to compare F-M's third reduced form with the form we would obtain from equation (10^*) or (10^{*a}) if we were to remove all autonomous elements from the left side of the equation.

We single out equations (10^*) and (10^{*a}) rather than ($10'$) and ($10'a$) because F-M explicitly wanted to work with money as one of the explanatory variables, and with the total (change of the) deficit included in the so-called autonomous expenditure, the other explanatory variable. This means that their equations (19a), (19b), and (19c) must be interpreted as reduced forms of models without any kind of built-in flexibility in budget and money supply in the wide sense. There is nothing wrong in that, as we have seen. The criticism raised against F-M by all participants of the discussion[15]—that they confuse autonomous and induced variables, particularly with respect to budget revenue and money—was thus misdirected. F-M were to that extent entirely justified in ridiculing this objection. Granted that, it should not be overlooked, however, that the effects that F-M tried to measure were the effects of total budget and money supply changes which lump together effects of discretionary measures and automatic effects of built-in flexibility and thus are of no relevance for appraising actual discretionary policy measures taken in the past and present, and (probably) in the future. F-M's findings are relevant for appraising policy actions only in the country of Erehwon; they would, however, become of great interest for evaluating past and present developments were they confronted with corresponding estimates of the effects of discretionary measures because they would then allow us to draw conclusions about the stabilizing or destabilizing properties of budget and monetary systems.

[15] Ando and Modigliani; Michael DePrano and Thomas Mayer (1965); Donald Hester.

Starting out, then, from equation (10^*) we obtain after deducting the autonomous part of dY,

$$(10^{*\prime})\quad dY-dc_3-di_3-dG = \frac{1}{\Delta^*}\Big[dc_3(1-\Delta^*)+di_3(\delta^*\mu m_1-\Delta^*) + dG(\delta^*\mu m_1-\Delta^*) - (dT-dB)(c_1+\delta^*\mu m_1-1) + \frac{dB}{r_g}(1+c_2-\delta^*(\mu m_1-m_3 i_1)) + D_0(\delta^*\mu m_1-1) + dM_2(1+c_2-\delta^*(\mu m_1+i_1))\Big]$$

This expression, compared with F-M's third reduced form (19c′), on differential form, gives rise to the following observations:

(i) Equation (10^*) is based on a model with the European type of budget constraint. It may be questioned to what extent this model is adequate for the *U.S.* economy after the Accord of 1951, and before the Great Depression. When dM_2 is exogenous, however, we had no other choice than to switch back to the European type of constraint. In a different model this might perhaps not be so. If the comparison had been between the effects of central bank money and budget changes on the basis of equation (10^{*a}), we could, and probably should, work with different models (different budget constraints) before and after 1951. This would lead to different specifications of the regression equations for different periods. F-M use the same specification for all periods.

(ii) Equation ($10^{*\prime}$) implies that autonomous expenditure, dA, should be defined as

$$dA = dc_3(1-\Delta^*) + (di_3 + dG)(\delta^*\mu m_1 - \Delta^*) - (dT - dB)(c_1 + \delta^*\mu m_1 - 1)$$

instead of

$$dA = d(I + G - T + B)$$

which corresponds to F-M's definition (disregarding the foreign deficit). The term A should thus be the weighted sum of the individual items. Only in the simplest Keynesian models where $\Delta^* = 1 - c_1$ can weighting be escaped (DePrano and Mayer). Since the weights of the individual items are composed of coefficients in the basic model and are unknown, we would have to use dc_3, $di_3 + dG$, and $dT - dB$, with D_0, dM_2, and dB/r_g as separate explanatory variables in a regression analysis. Similar points were made by Ando and Modigliani.

(iii) The exclusion of the autonomous parts of dY from the left-hand side of ($10^{*\prime}$) implies that the multiplier effects dc_3, di_3, and dG are reduced by one, whereas those of $dT - dB$, dB/r_g, D_0, and dM_2 remain unchanged. This was, of course, pointed out by F-M (1963, pp. 176 and 205); but both they and other participants in the discussion seem to have overlooked that whereas $1/\Delta^*$, the general multiplier, has to be positive for the system to be stable (if this is thought necessary), there is nothing to prevent $1/\Delta^* - 1$, the consumption multiplier (of dc_3 in our model) from being negative. F-M found negative consumption multipliers and correlation coefficients between autonomous expenditure and consumption for a number of periods and declared this to be "the most puzzling result we have found" (1963, pp. 203 and 212). Nothing could be less puzzling; it only requires that $1/\Delta^* < 1$, that is $\Delta^* > 1$. This suffices to make the effects of autonomous consumption, dc_3, negative. And the effects of $(di_3 + dG)$ and $-(dT - dB)$ are in any case negative with our model. With the specifications of our model, F-M should indeed expect to find negative multipliers for what they call autonomous expenditure.[16] And apart from the simplest Keynesian models, the consumption multiplier may always be negative.

On the other hand, it cannot be sufficiently emphasized that for the very same economy which gave us negative effects of total budgetary changes, we should find positive effects of discretionary fiscal measures, equation (10′) and they might even be larger, dollar by dollar, than the effects of discretionary monetary measures. If this were what the Keynesians have in mind, both parties could be perfectly right, asking different questions to the same set of economic relations.

(iv) We are led to consider the numerical size of Δ^*. Our assumptions about

[16] Also the St. Louis econometricians, Leonall Anderson and Jerry Jordan (1968) (1969) find systematically small positive or negative effects (usually insignificant) effects on *GNP* of their fiscal indicators with strong and significant positive effects of money supply in the wide sense. Their fiscal indicators are, however, open to criticism. It is also interesting to note that when corresponding regressions are made with the rate of interest as the monetary policy variable, the crowding effects of government spending seem to disappear; fiscal policy then appears with strong effects (Spencer and Yohe, p. 23; M. Keran (1970)). This is precisely what we should expect with our model when we give up M_2 as exogenous variable and uses $\bar{r}$ instead.

Anderson and Jordan work with distributed lags in their reduced forms and their statistical estimates show invariably that fiscal policy has positive effects on Y in the first period(s) but that effects thereafter become strongly negative. By lagging all monetary variables in the behaviorial equations of our model, experiments show that our model on an appropriate choice of coefficient values can reproduce exactly the same time pattern of the response of income to fiscal policy changes: positive effects in the first period(s), followed by stronger negative effects. From our discussion of the individual terms in equation (10*) it is easy to understand why this is so. It is the repercussions through the financial part of the system that makes the ultimate results differ from standard Keynesian thinking, and if these repercussions are lagged the Keynesian relations will dominate in the beginning but sooner or later be wiped out by the financial reactions.

the signs of the constants and parameters do not guarantee $\Delta^*>0$, but do not prevent $\Delta^*>1$, either. If m_1 is numerically small, this condition may be fulfilled; if, further, c_2, B, i_1, i_2, or μ are sufficiently large (numerically) we will indeed have $\Delta^*>1$.

(v) Note also that in regression analysis of the reduced forms we would probably have to use the solution on the form of (10*′), where Terms II and III of (10*) are consolidated. Otherwise, we might run the risk of a high degree of collinearity between the explanatory variables. For empirical analysis we might thus not even have the option of excluding the so-called monetary effects of budgetary changes from the effects of total fiscal changes to escape the negative effects of the latter.

(vi) Given the general multiplier, $1/\Delta^*$, we know that (with Terms II and III of (10*) consolidated) the effects of budget changes are negative and those of changes in money supply positive. Inspection of the weights in (10*′) shows, moreover, that both are numerically larger, the larger are c_2, i_1, and m_1 (numerically). In other words, large wealth and interest responses increase the effects of money supply as compared with the (negative) budgetary effects. By insisting on large wealth and interest responses, Keynesians may thus be doing harm to their own case, provided that the issue is the total effects. With respect to the effects of discretionary measures, however, this does not hold true.

(vii) Finally, the specification of F-M's third hypothesis is open to objections, quite apart from the definition of dA. First, debt management and open market operations implying changes in the privately owned stock of government bonds are not included in their regressions. This is a misspecification; only if they had chosen to let all bond transactions appear as endogenously determined would it have been correct to leave out government bonds in the regression equation; but in that case we would have had to let the discount rate appear as an exogenous variable and a determinant in the regressions.

(viii) Second, our model leads to the inclusion of the initial budget deficit as an explanatory variable in the differential model, in addition to the (weighted) changes in the budget. F-M included the budget deficit in their original equations and on differentiation only the change in the deficit should appear as an explanatory variable. There is no doubt that F-M at this point have been perfectly faithful to conventional Keynesian theory; the appearance of the deficit itself is at variance with basic Keynesian thinking, excepting Lerner; and is closely related to the recognition of the public sector budget constraint. Models that respect that constraint on public sector behavior will always come out with the budget deficit itself as an explanatory variable when working on first differences.

REFERENCES

G. **Akerlof**, "The Demand for Money: A General Equilibrium Inventory Theoretic Approach," *Rev. Econ. Stud.*, Jan. 1973, *40*, 115–30.

L. C. **Andersen** and J. L. **Jordan**, "Monetary and Fiscal Actions: Test of their Relative Importance in Economic Stabilization," *Fed. Reserve Bank St. Louis Rev.*, Nov. 1968, *50*, 11–23.

——— and ———, "Reply," *Fed. Reserve Bank St. Louis Rev.*, Apr. 1969, *51*, 12–16.

A. **Ando** and F. **Modigliani**, "The Relative Stability of Monetary Velocity and the Investment Multiplier," *Amer. Econ. Rev.*, Sept. 1965, *55*, 693–728.

K. **Brunner**, "A Schema for the Supply Theory of Money," *Int. Econ. Rev.*, Jan. 1961, *2*, 79–109.

———, "Monetary Analysis and Federal Reserve Policy," in K. Brunner, ed., *Targets and Indicators of Monetary Policy,* San Francisco 1969, 250–82.

——— and **A. H. Meltzer**, "The Meaning of Monetary Indicators," in G. Horwich, ed., *Monetary Process and Policy: A Symposium,* Homewood 1967, 187–217.

———, "A Monetarist Framework for Aggregative Analysis," Proc. First Konstanz Conference on Monetary Theory and Policy, *Kredit und Kapital,* 1972.

C. F. Christ, "A Short-Run Aggregate Demand Model of the Interdependence and Effects of Monetary and Fiscal Policies with Keynesian and Classical Interest Elasticities," *Amer. Econ. Rev. Proc.,* May 1957, *57,* 434–43.

———, "A Simple Macroeconomic Model with a Government Budget Restraint," *J. Polit. Econ.,* Jan. 1968, *76,* 53–67.

M. DePrano, and T. Mayer, "Tests of the Relative Importance of Autonomous Expenditure and Money," *Amer. Econ. Rev.,* Sept. 1965, *55,* 729–52.

A. C. Enthoven, "A Neo-Classical Model of Money, Debt, and Economic Growth," Mathematical Appendix in J. G. Gurley and E. J. Shaw, *Money in a Theory of Finance,* Washington, 303–359.

L. L. Gramley, and S. B. Chase, Jr., "Time Deposits in Monetary Analysis," *Fed. Res. Bull.,* Oct. 1965, *51,* 1380–1406.

M. Friedman, "A Theoretical Framework for Monetary Analysis," *J. Polit. Econ.,* Mar./Apr. 1970, *78,* 193–238.

——— and **D. Meiselman**, "The Relative Stability of Monetary Velocity and the Investment Multiplier in the United States, 1897–1958," in *Stabilization Policies: Commission on Money and Credit,* Englewood Cliffs, 1963, pp. 165–268.

——— and ———, "Reply to Ando and Modigliani and DePrano and Thomas Mayer," *Amer. Econ. Rev.,* Sept. 1965, *55,* 753–85.

P. Frost, "Discussion," *Amer. Econ. Rev. Proc.,* May 1970, *66,* 55–6.

M. J. Hamburger, "Indicators of Monetary Policy: The Arguments and the Evidence," *Amer. Econ. Rev. Proc.,* May 1970, *66,* 32–39.

B. Hansen, *Finanspolitikens ekonomiska teori,* SOU 1955:25 Stockholm, 1955; English translation: *The Economic Theory of Fiscal Policy,* London 1958; Lund 1967.

———, *Fiscal Policy in Seven Countries,* Paris 1969.

D. D. Hester, "Keynes and the Quantity Theory," *Rev. Econ. Statist.,* Nov. 1964, *46,* 364–68.

M. W. Keran, "Monetary and Fiscal Influences on Economic Activity—The Historical Evidence," *Fed. Reserve Bank St. Louis Rev.,* Nov. 1969, *51,* 5–23.

———, "Selecting a Monetary Indicator—Evidence from the United States and Other Developed Countries," *Fed. Reserve Bank St. Louis Rev.,* Sept. 1970, *52,* 8–19.

A. P. Lerner, "Functional Finance and the Federal Debt," *Soc. Res.,* Feb. 1943, *10,* 38–51.

———, *The Economics of Control,* New York 1946.

L. A. Metzler, "Wealth, Saving, and the Rate of Interest," *J. Polit. Econ.,* Apr. 1951, *59,* 93–116.

R. A. Musgrave, *The Theory of Public Finance,* New York 1959.

———, "On Measuring Fiscal Performance," *Rev. Econ. Statist.,* May 1964, *46,* 213–20.

B. Ohlin, *Penningpolitik, offentliga arbeten, subventioner och tullar som medel mot arbetslöshet, Bidrag till expansionens teori,* SOU 1934:12 Stockholm 1934.

———, *Kapitalmarknad och räntepolitik,* Stockholm 1941.

K. Popper, *The Logic of Scientific Discovery,* New York 1959.

P. A. Samuelson, "Simple Mathematics of Income Determination," in L. A. Metzler et al. eds., *Essays in Honor of Alvin H. Hansen,* New York 1948, 133–55.

W. L. Silber, "Fiscal Policy in IS-LM Analysis: A Correction," *J. Money, Credit, Banking,* Nov. 1970, *2,* 461–73.

R. W. Spencer and W. P. Yohe, "The 'Crowding Out' of Private Expenditures by

Fiscal Policy Actions," *Fed. Reserve Bank St. Louis Rev.*, 1970, *52*, 12–24.

J. Tobin, "Monetary Semantics," in K. Brunner, ed., *Targets and Indicators of Monetary Policy*, San Francisco 1969, 165–74.

K. Wicksell, *Finanztheoretische Untersuchungen*, Jena 1898.

Board of Governors of the Federal Reserve System, *The Federal Reserve System*, Washington 1963.

International Monetary Fund, *International Financial Statistics*, Washington, various years.

A Model of the Trade and Government Sectors in Colonial Economies

By THOMAS BIRNBERG AND STEPHEN RESNICK*

This study examines historically and quantitatively the process of colonial development for selected economies of Asia, Africa, and Latin America. A simultaneous equations model is applied to each country from about the start of the twentieth century until the outbreak of World War II in an attempt to identify those forces which made these economies dependent on international trade. A principal objective of this research is to construct and apply a structural model of aggregate behavior which explains the development pattern of several countries in different geographical areas. The sample countries include those experiencing overt colonial control such as Ceylon, India, Jamaica, Nigeria, Philippines, and Taiwan, as well as Cuba, Chile, Egypt, and Thailand where foreign influence and control were perhaps more subtle but no less important in determining economic activity. Our analysis does not intend nor does it show the actual degree or level of exploitation or which groups within which countries benefitted or lost from the colonial relationships.

From about the end of the nineteenth century onwards, economic development in these countries took on a new direction with the extensive penetration of Western commodities, organization, and control, ushering in the era of the export economy.[1] The model measures these external forces through such variables as real income and prices within the developed world which shifted the export demand function facing the colony. This historical development was also characterized by internal colonial government expenditures, directed towards the promotion of the export economy. The effect was a rightward shift of the colonial export supply function.[2] Colonial export prices are thus shown to be endogenous and determined simultaneously by expenditures of the colonial government and also by the developed countries' demand characteristics and trade policies. This suggests that previous explanations of colonial development which have omitted the government as an explicit variable have been misspecified.

The paper is divided into the following three sections: Section I presents the econometric model and summarizes the estimation method used; the empirical results are presented and discussed in Section II; and the last section evaluates the results using dynamic simulations.

* Assistant professor, Yale University and professor, University of Massachusetts, respectively. Portions of this research were financed by funds provided by the National Science Foundation, GS-2804. However, the views expressed in this paper do not necessarily reflect those of the Foundation. Ignazio Cocco, Janet Farooq, and Robin Kibuka provided valuable research assistance in the preparation of this paper.

[1] We selected countries for which we could collect reliable and consistent data going at least as far back as the start of the twentieth century. The sample period for each country is as follows: Ceylon, 1897–1938; Chile, 1890–1938; Cuba, 1903–1937; Egypt, 1891–1937; India, 1890–1936; Jamaica, 1886–1938; Nigeria, 1901–1937; Philippines, 1902–1938; Taiwan, 1904–1936; and Thailand, 1902–1936. A data set for each country was collected from original country and colonial sources listed in the bibliography and is available on request from the authors. Considerable effort was devoted to constructing the data set in order to have a consistent time-series not only to insure reliable estimates but also to make intercountry comparisons. Several countries which could have been included in our sample were rejected because of the quality of their original data.

[2] For a theoretical model describing the role of the government sector, see Stephen Hymer and Resnick.

I. The Model

A basic circular structure underlies a system of five behavioral equations, specified in double logarithmic form,[3] and five definitional equations which together describe the trade and government sectors of a colonial economy. Colonial government expenditures are directed toward promoting the growth of real exports which, in turn, pay for real imports. The expansion of exports and imports generate directly and indirectly revenues for further government expenditures which continue the growth process and complete the circular structure of the model. Changes in real income, prices, and trade policies within the developed world affect the colonial structure through the developed countries' demand for raw materials and food.

A. *The Trade Sector*

Supply and demand equations for real exports along with a market clearing equation act together to describe the economic interrelationship between colony and colonizer. The trade sector is completed by a colonial demand equation for real imports and by definitional equations for nominal exports and imports.

Equation (1) determines the principal commercial activity of the colonial economy which is the supply of real exports.[4]

$$(1)\quad \ln X_{S_t}^R = a_0 + a_1 \ln Px_t + a_2 \ln Pm_t + a_3 \ln \sum_{i=1}^{\infty} G_{t-i}^R + a_4 \ln X_{S_{t-1}}^R + a_5 D_{S_t}$$

where:

X_S^R = the supply of total real commodity exports from the colony.

Px = the colony's Paasche export price index (1913 = 1).[5]

Pm = the colony's Paasche import price index (1913 = 1).

$\Sigma_{i=1}^{\infty} G_{t-i}^R$ = the lagged value of accumulated real government expenditures in the colony.[6] In the discussion to follow, this variable will be denoted *ARGE*.

D_S = a dummy variable reflecting the impact of exogenous events upon the colony's supply function.

This aggregate supply equation assumes that a composite commodity, real exports, depends upon a corresponding export price index. A priori the sign of the export price coefficient a_1 is expected to be positive.[7] Import prices influence real exports in two ways: the costs of production are assumed to be represented by this index since these economies were dependent on the interna-

[3] The double logarithmic form will provide a linear linkage between the model's real and nominal variables, and the estimated coefficients will be elasticities. Linear specification was tried and was found to yield highly inferior estimates, because of multiplicative rather than additive errors.

[4] For ease of exposition, the error terms in all the behavioral equations in this section have been omitted.

[5] The Paasche export and import price indices were specially calculated for this study. These indices were calculated using the largest bundles of goods for which consistent and reliable quantity and value data were available. Paasche indices rather than Laspeyres or Fisher Ideal indices were used because the composition of the commodity bundles changed more rapidly than did prices. The indices were linked using backward bases because exporters respond to current prices relative to past prices, not relative to future prices. The linkages of the indices were designed to account for the principal changes in the composition of the commodity bundles.

[6] The method of calculating this variable is found in the description of the government sector.

[7] a_1 is not only the export price elasticity, but also the terms of trade elasticity, for the supply equation can be rewritten as

$$(1')\quad \ln X_{S_t}^R = a_0 + a_1 \ln P_{T_t} + (a_2 - a_1) \ln Pm_t + a_3 \ln \sum_{i=1}^{\infty} G_{t-i}^R + a_4 \ln X_{S_{t-1}}^R + a_5 D_{S_t}$$

where: $P_T = Px/Pm$ = colony's terms of trade.

tional market for many of their intermediate goods and almost all of their capital goods;[8] and the domestic price of incentive or wage goods in the export sector is assumed to be reflected by this index since their importation often led to the displacement of inferior rural manufactures by superior foreign commodities.[9] For these reasons the sign of the import price coefficient is expected to be negative. Taken together, the export and import prices reflect the macroprofitability of an export economy.

Accumulated real government expenditures enter as a variable because the government is assumed to have been a crucial provider of the necessary infrastructure and social intermediate products associated with the development of an export economy. The growth of trade experienced by these economies would hardly have been possible without the expenditures on harbors, wharves, culverts, road systems, railroads and other public works as well as the investments in administrative infrastructure, in health facilities such as malaria control, in the establishment of order such as an organized police and army establishment, and in various directly productive agrarian activities such as irrigation, artesian wells, disease research for crops, and communication facilities.[10] Thus, accumulated real government expenditures provided the necessary colonial environment in which producers were able to respond to changing market incentives reflected by export and import prices. In a real sense colonial technical progress was embodied within government expenditures, thereby providing the favorable "atmosphere" for the historic development of the export economy.[11] A priori the expected sign of the government variable is positive indicating a rightward shift over time in the supply schedule.[12] And the actual size of the coefficient provides empirical information on the marginal productivity of colonial government expenditures in promoting real exports.

We rejected time as a variable in our model by testing whether lagged accumulated real government expenditures,

$$ARGE \equiv \sum_{i=1}^{\infty} G_{t-i}^{R}$$

was a time variate in the supply equation. First, we replaced *ARGE* with time, and the results contained incorrect signs and high standard errors. Then, we included both *ARGE* and time, and the time variate was insignificant while *ARGE* remained significant. The main reason that *ARGE* is not a proxy for time is because it does not grow linearly. In particular there exist many substantial differences between the growth rates of G_t^R for various subperiods of the estimation period.

Lagged real exports were introduced into equation (1) to test for the possibility that there was a difference between the short- and long-run supply response. The supply equation was estimated with and

[8] One might reasonably assume that domestic labor was available in unlimited supply for export production.

[9] See Resnick (1970) for such a model of trade behavior.

[10] Government expenditures are not disaggregated by category basically because such a breakdown was available only for a few countries. Even if a consistent breakdown was available for all the colonies, there is no theory suggesting what was an export development expenditure and what was not. The only exception to this approach was to subtract expenditures neither directly nor indirectly connected with the development of an export economy, such as the royal household expenditures in Thailand.

[11] Private investment was omitted from the model not only because data were not available, but also because historically government expenditures were complements to or necessarily preceded private investment. In this sense, government expenditures provided the "big push" for the development of the export economy.

[12] We explored the specific question of the time pattern of governmental impact by estimating equation (1) using alternative distributed lag structures. These estimates yielded results inferior to those derived using *ARGE*.

without X^R_{t-1}, and we leave this choice of specification as an empirical question.

Equation (2) explains the demand for real exports.[13]

$$(2)\quad ln\, X^R_{D_t} = b_0 + b_1\, ln\, Px_t + b_2\, ln\, Y^R_t + b_3\, ln\, Pd_t + b_4\, ln\, X^R_{D_{t-1}} + b_5 D_{D_t}$$

where:

- X^R_D = the demand for total real commodity exports.
- Y^R = real *GNP* in the developed country; for some countries industrial production, Q, is used.
- Pd = the domestic price level in the developed country; when Y^R appears as a variable, Pd is the implicit *GNP* price deflator; when Q is used, a Pd was appropriately selected to reflect the commodities traded as either an import price index, or a price index of raw materials.
- D_D = a dummy variable reflecting the effect of exogenous events on the developed country's demand function.

As with the supply equation, we specify a standard distributed lag formulation.

The colony is not assumed to be a price taker; rather the market clearing equation (3) holds.

$$(3)\qquad ln\, X^R_{S_t} = ln\, X^R_{D_t}$$

Export prices are then endogenous to the colonial trade system. This assumption is in contrast to much of the development literature where export prices are assumed implicitly to be exogenous to the developing world.

The assumption of an exogenous export price was empirically tested by estimating the reduced form equation for the export price. Exogenous variables in both the supply and demand equations for each country were significant and, therefore, were required to explain the export price. Thus the assumption of an exogenous export price was rejected,[14] and the estimation of a demand equation as well as a supply equation using simultaneous equations methods was necessary.

The political and economic relationship between the colony and the developed country often led to a fairly high percentage of the former's total commodity trade being carried on with the developed country and its other colonies (see Table 1). Exports to these other colonies were either for transshipment (for example, entrepôt trade with Hong Kong and Singapore) or for direct consumption in these colonies. The model assumes that the economic activity of these other colonies is reflected by and can be measured by that of the developed country. Therefore, the specification of equation (2) is based upon the empirical observation that colonial trade was bilateral in nature which, in turn, reflected the bilateral political relationship that emerged over time. For these reasons, competitive prices for alternative sources of export supply do not appear in the developed country's aggregate demand schedule. A domestic price level is included to reflect the substitution between domestic and imported goods.

The incomes and prices of the United Kingdom, United States, and Japan are assumed to be the main driving forces or instruments affecting the economic ac-

[13] For countries with a variable exchange rate, the demand price is a new variable Px' defined by the additional equation, $ln\, Px' = ln\, Px + ln\, \pi$, where π is the exchange rate of the colony's currency relative to that of the developed country to which it was tied.

[14] The importance of specifying the export price as endogenous was originally motivated by estimates of the supply equation with the export price as exogenous. These estimates yielded negative, rather than positive, export price coefficients for most countries.

TABLE 1—DISTRIBUTION OF EXPORTS FOR SELECTED YEARS

	Developed Country	Percentage of Export Trade with Developed Country 1900[a]	1913	1925	1938[b]	Instruments Corresponding to Developed Country[c]
Ceylon	*U.K.*+colonies	78.3	58.0	52.2	72.1	Y^R_{UK}, *LIMIT*
	U.S.	7.0	16.5	29.9	12.6	*CAR*
Chile	*U.S.*	3.9	21.0	39.2	30.0	Q_{US}, P^{mn}_{US}, *RESTR*
	U.K.	73.5	39.5	34.6	26.0	
Cuba	*U.S.*	76.8	79.9	74.6	80.7	Q_{US}, P^m_{US}, *QUOTA*
Egypt	*U.K.*	54.5	42.6	43.5	30.9	Y^R_{UK}, P_{UK}, *WWI*
India	*U.K.*+colonies	59.1	39.4	34.1	45.7	Q_{UK}, P_{UK}
Jamaica	*U.K.*+colonies	23.2	24.8	48.6	85.0	Q_{UK}, P_{UK}
	U.S.	63.8	57.4	40.7	3.7	*RESTR*
Nigeria	*U.K.*	31.4	50.9	54.7	50.0	Y^R_{UK}, P_{UK}
	Germany	55.8	41.8	21.3	17.7	*WWI*
Philippines	*U.S.*	40.0	34.4	73.0	77.0	Y^R_{US}, P_{US}, *QUOTA*
Taiwan	Japan	60.0	75.7	82.0	89.4	Y^R_{JP}, P_{JP}, *WWI*
Thailand	*U.K.*+colonies	85.5	83.2	74.6	79.0	Y^R_{UK}, P_{UK}

[a] The initial year used for Cuba is 1903, for Nigeria is 1901, for Philippines is 1904, and for Thailand is 1901.
[b] The final year used for Chile, Cuba, and Egypt, 1937; for India, 1936; and for Taiwan, 1935.
[c] For definitions see Appendix.

tivity of their respective colonies (see Table 1). While the model focuses upon the principal trading relationships that emerged historically, adjustments to this approach are made when the export trade pattern showed an obvious change in direction. For Ceylon, the variable *CAR* reflects the demand for rubber by the United States after 1905. In Chile, even though the United States and the United Kingdom were the major trading partners, copper exports were controlled by three *U.S.* companies. The *U.S.* income and price variables are taken then as the main determinants of the demand function. For Jamaica, the dummy *RESTR* accounts for the dramatic decline of its trade with the United States after the imposition of the Hawley-Smoot Tariff of 1930. For Nigeria, the dummy *WWI* measures the impact of the complete cutoff of its trade with Germany.

The growth in real exports had as its dual the growth of real imports. The increased specialization of the colonial economy was reflected by a shift of resources out of traditional activities into commercial ones. Correspondingly, the demand for foreign consumer and intermediate commodities expanded. Here, the opposite of an import-substitution policy was being pursued. Colonial policy was clearly biased towards the promotion of exports rather than indigenous manufacturing, and the resulting decline of traditional industry associated with the precolonial agrarian society was replaced by the expansion of and reliance on imported manufactures.

Equation (4) describes the demand function for imports in real terms.

$$(4) \qquad ln\, M_t^R = c_1\, ln\, X_t^R + c_2\, ln\, Pm_t + c_3\, ln\, Px_t + c_4 D_{M_t}$$

where:

M^R = total real commodity imports by the colony.

D_M = a dummy variable reflecting the impact of exogenous events upon the colony's import function.

The following definitional equations for nominal commodity imports M_t and nominal commodity exports X_t complete the trade sector.

$$(5)\qquad ln\ M_t = ln\ M_t^R + ln\ Pm_t$$

$$(6)\qquad ln\ X_t = ln\ X_t^R + ln\ Px_t$$

In equation (4) the coefficient c_1 measures the increase in real imports caused by an increase in real exports. This coefficient also provides a measure of the nominal trade balance. If $c_1 < 1$, then the colony ran a nominal trade surplus, while if $c_1 > 1$, it ran a nominal trade deficit.[15] The coefficient c_2 is the import price elasticity of demand by the colony for developed countries' goods. The coefficient c_3 measures the shift of the demand schedule for real imports as export prices change. A priori, we expect the sign of c_2 to be negative and the sign of c_3 to be positive.[16]

B. *The Government Sector*

The government sector has two behavioral equations, the generation of nominal revenues (7) and the expenditures from that revenue (8).

$$(7)\quad ln\ R_t = d_0 + d_1\ ln\ X_t^R + d_2\ ln\ M_t + d_3 D_{R_t}$$

$$(8)\quad ln\ G_t = e_1\ ln\ R_t + e_2\ ln\ G_{t-1} + e_3 D_{G_t}$$

where:

R = total nominal revenues generated in the colony.
M = the colony's total nominal commodity imports.
G = nominal colonial government expenditures.
D_R; D_G = dummy variables reflecting the impact of exogenous events upon the colony's government revenue and expenditure functions, respectively.

[15] To derive this conclusion, rewrite equation (4) in terms of nominal imports and nominal exports as

$$(4')\qquad ln\ M_t = c_1\ ln\ X_t + (c_2 + 1)\ ln\ Pm_t + (c_3 - c_1)\ ln\ Px_t + c_4 D_{M_t}$$

[16] With both the import and export prices scaled equal to one in 1913, the coefficients c_2 and c_3 describe how these prices change the nominal trade balance of the colony relative to that in 1913.

The expansion of real exports is assumed to generate revenues directly from specific export taxes or indirectly given that taxable real economic activity in the colony was in one way or another tied to real export activity. Revenues from nominal imports were derived directly from import duties and indirectly from taxes on commercial import activity. The expenditure equation (8) can be derived from a revenue expectation model, where government spending in the current period depends on expected revenue. An alternative derivation of equation (8) is that government expenditures are divided between recurrent expenditures, $e_2\ ln\ G_{t-1}$, and current expenditures, $e_1\ ln\ R$.

Real government expenditure in a given year is defined as[17]

$$(9)\qquad ln\ G_t^R = ln\ G_t - ln\ Pm_t$$

The price of imported goods is thus assumed to reflect the cost of government expenditures because government capital goods were usually imported, and government employees, particularly colonial officers, were dependent on imports to maintain their standard of living. Lagged accumulated real government expenditures (*ARGE*) are then calculated using the inventory formula[18]

$$(10)\quad ARGE \equiv \sum_{i=1}^{\infty} G_{t-i}^R = \sum_{i=1}^{\infty} G_{T_0-i}^R + \sum_{i=T_0}^{t-1} G_i^R$$

[17] For some countries, we added to government expenditures in equation (9) the relatively small value of public works expenditures financed from borrowing abroad.

[18] For several reasons, accumulated real government expenditures were not depreciated. There exists no a priori information on the actual depreciation of these expenditures. If they are a measure of colonial technical progress, then no depreciation rate should be used. In addition, we carefully checked the sensitivity of our supply equation estimates against those with assumed rates of depreciation. For a range of possible rates of 1 to 4 percent, we found small changes in the estimates. As the depreciation rate rose above 4 percent, the estimates became increasingly inferior.

where: T_0=earliest year of data for G_t^R. $\sum_{i=1}^{\infty} G_{T_0-i}^R$ is the initial stock of accumulated real government expenditure which was calculated by first estimating the regression equation for the growth rate of G^R:

$$\ln G_t^R = a_0 + a_1 t \qquad t = T_0, T_0+1, \ldots, \tau$$

where τ is selected as the year with the longest consistent pattern of growth of G^R. Let $\hat{a}_0$ and $\hat{a}_1$ be the estimated values of a_0 and a_1. Then

$$\sum_{i=1}^{\infty} G_{T_0-i}^R = \sum_{i=1}^{\infty} e^{\hat{a}_0} e^{(T_0-i)\hat{a}_1} = e^{\hat{a}_0+T_0\hat{a}_1} \sum_{i=1}^{\infty} e^{-\hat{a}_1 i}$$
$$= e^{\hat{a}_0+T_0\hat{a}_1}/(e^{\hat{a}_1} - 1)$$

is the estimated initial stock of accumulated real government expenditures.

C. *Estimation Method*

Equations (1) through (10) constitute a system of ten equations in ten unknowns, namely X_S^R, X_D^R, Px, M^R, X, M, R, G, G^R, and $ARGE$. Initial estimates of the model using instrumental variables revealed a well-defined pattern of serial correlation in the plotted residuals of most equations. This serial correlation posed serious problems for making reliable cross-country comparisons, for distinguishing between distributed lags and autoregressive errors, and for judging the performance of the model in explaining the observed development pattern.

An estimation method described in the econometric part of the Appendix was used to solve these problems of autocorrelated errors in a simultaneous equation system. All current and lagged endogenous variables in the autoregressive model were treated as endogenous. Therefore, these variables were instrumentally adjusted taking into account the structural ordering of the model. Using an iterative method, the model yielded consistent estimates of the equation parameters. This method always converged to a value of the autoregressive parameter ρ that satisfied the condition that its absolute value did not exceed one. The method actually removed the first-order serial correlation pattern in the data.[19] Finally, the iterates yielded estimates of the parameters with correct signs and relatively low standard errors.

II. The Results

The estimates of the five structural equations for each of the ten countries are reported in Appendix Tables 1 through 5. All the signs of the estimated coefficients are correct, and the standard errors of the coefficients and of the equations are rather low.[20]

We find lagged real exports X_{t-1}^R appearing as an explanatory variable indicating the presence of a lagged adjustment process in the supply equations of Cuba, Ceylon, Nigeria, and Taiwan, and in the demand equations of Chile, Nigeria, and the Philippines. For the remaining supply and demand equations, lagged real exports did not appear significant and it was dropped as an explanatory variable.

The estimated export price coefficients of the supply equations indicate inelastic aggregate supply schedules in colonial economies. The elasticity estimates vary in magnitude from .142 in Nigeria to .536 in Jamaica, with an average of .315. These are short-run supply elasticities for all countries and also long-run supply elasticities for the six countries where no dis-

[19] While this conclusion is based upon examination of the plots of the estimated residuals of the fifty equations estimated, one imperfect indicator of the degree of improvement is given by the Durbin-Watson statistic. Omitting the eight equations for which no autoregressive adjustment was needed, the average $D.W.$ statistic was 1.23 before and 1.94 after the adjustment.

[20] We have reported t-values and the R^2 statistics because they are usually expected, even though there are qualifications about their applicability in a simultaneous equation system.

tributive lag process was found. For the four countries in which X^R_{t-1} appears, the average long-run elasticity is .880 ranging from .281 in Nigeria to 1.92 in Ceylon. If Ceylon is excluded on the grounds that its exports were comprised primarily of output from tea and rubber estates, then the estimated aggregate supply elasticity for a colonial economy is inelastic in both the short and long run.[21]

The elasticity of demand for aggregate real exports is also typically inelastic with the average export price coefficient in the demand equation of $-.512$ in the short run for all ten countries. The long-run elasticity averages -1.15 for three countries in which an adjustment process is present. If Chile is excluded on the grounds that the demand for the two main exports, copper and nitrate, is elastic, then the average falls to $-.41$ in the short run and $-.95$ in the long run (for Nigeria and the Philippines).

The relatively inelastic supply and demand schedules for a colonial economy when both curves were quite shiftable produced rather dramatic changes in the price of exports. Shifts in the supply schedule are measured by the coefficients of the import price and accumulated real government expenditures. The average coefficient associated with the government variable is .47 and with the import price, $-.35$; for the four countries with an adjustment process, the respective long-run coefficients are 1.2 and $-.8$. India has the lowest long-run elasticity associated with government expenditures followed by Egypt and then Thailand. In fact, these three countries seem to be in a relatively low export-promoting governmental productivity group compared to the other seven countries; the average long-run governmental elasticity being 1.00 for the latter group and .31 for the former. Almost one-half of the Indian budget was devoted on the average to military expenditures for the period and thus a low coefficient is not surprising. Although Egypt and Thailand were not colonies in the legal sense, their respective economies were as much subject to *U.K.* influence as that of India. Egypt had the slowest growth rate of real exports of all ten countries and her resources seem to have been increasingly focused on the required repayment of previous international loans rather than devoted toward development expenditures.[22] The Thai government was effectively constrained from controlling and utilizing governmental expenditures for productive investments by *U.K.* financial control.[23]

This result shows that the government must be included in explaining the development of a colonial export economy and that to ignore or omit this relationship would lead to a major error of misspecification. These conclusions do not depend upon the size of the coefficient of accumulated real government expenditures. The impact of these expenditures is equally important in explaining the low export growth of India which had the lowest coefficient as well as in explaining the high export growth of Ceylon and the Philippines which had the highest coefficients.

Shifts in the demand schedule facing the colony are measured by the coefficient of real income and domestic prices in the developed country. The average income elasticity of demand for colonial goods is .83 in the short run and 1.39 in the long run for those three countries in which X^R_{t-1} appears. For all ten countries the average long-run income elasticity is 1.09 which indicates the importance of income growth in the developed countries on ex-

[21] Of course, within the export sector, there may be significant shifting of resources out of one export to another as relative export prices change.

[22] See Charles Issawi and A. E. Crouchley for further discussion.

[23] See James Ingram.

port growth in the colonial countries. The average elasticity of substitution between home and colonial commodities in the developed countries is .45 in the short run and, for the latter three countries, 1.33 in the long run. Examining these substitution elasticities by colonial blocs suggests that those countries which were under direct or indirect United States influence had the highest substitution effects as compared to the United Kingdom bloc or to Japan (for the case of Taiwan). There was more internal substitution over the period within the United States (copper for Chile and cane sugar for Cuba and the Philippines) than for either the United Kingdom or Japan. The model would predict, then, that *ceteris paribus*, a fall in *U.S.* prices would shift its demand schedule for colonial goods to the left more than an equivalent fall in the prices of the United Kingdom or Japan. For this reason the world depression of the 1930's had a more dramatic effect on Chile, Cuba, and the Philippines compared to the other seven countries. In contrast, the real income effects of the United Kingdom and Japan were more important than their price levels in influencing real export activity in their trade dependent countries.

The remaining variables in the supply and demand schedules are the dummy variables, which are specific to the economic history of the country such as the impact of an imposed tariff, quota, restriction scheme, or the influence of the First World War. The dummies all have the proper sign and are generally important in magnitude. The depressing effects of the restrictive trade policies pursued by the United States during the 1930's are measured by the dummies *RESTR* for Chile, and *QUOTA* for both Cuba and the Philippines. Since prices and real income were falling in the United States during the Great Depression, the imposition of these restrictions on trade compounded the difficulties experienced in these three countries as compared to countries tied to the United Kingdom or Japan. Interestingly enough, the trade restrictions put upon the Philippines were not as severe and came somewhat later in the depression years as compared to those for Cuba and Chile. This partial evidence suggests, at least for those countries linked to the United States, that legal colonialism as exemplified by the Philippines acted to mitigate the impact of United States policies during the 1930's.

The results indicate that similar dummy variables used in different countries need not have the same sign. For example, the effects of World War I benefitted the supply of real exports from Taiwan to Japan, whereas Nigerian trade suffered from the effects of the war because trade with Germany, one of its main export markets, was completely cut off and did not resume until 1922.

In the import equation, the average coefficient of real exports was .98, which indicates that, *ceteris paribus*, the growth of real exports in each colony did produce a corresponding demand for imported commodities. Moreover, this average coefficient indicates that the colonies ran a nominal trade surplus over this historical period which agrees with their actual balance of trade data. Jamaica was the only country with a coefficient greater than one, and Thailand's coefficient is the lowest of all ten countries. Both of these results are consistent with their actual trade data which show Jamaica as the only country with a nominal trade deficit and Thailand with the relatively largest nominal trade surplus. Finally, the average import price elasticity of demand was $-.42$, thus once again revealing inelastic demand schedules but this time for imports of the developed countries' goods.

The revenue equation shows that the average contribution of real exports to

nominal revenues was slightly higher (.59) than that of nominal imports (.52). However, the marginal contribution of real exports varies from a low of .230 in Thailand to a high of 1.203 in Chile (where imports were not effectively taxed), while the marginal contribution of nominal imports varies from a low of .191 in Cuba to .626 in Jamaica.

A measure of the total tax effort is the sum of the coefficients of X^R and M in the revenue equation. The average sum for the ten countries is 1.01 which suggests a constant returns to scale revenue function. However, Thailand and Taiwan have an average sum of tax coefficients of .61. Compared to the other eight countries where the average sum is now 1.12, the colonial tax effort for these two countries was not strong enough.

Some of the dummy variables in the revenue equation measure changes in the tax effort such as the imposition of an income tax on copper in Chile from 1926 to 1938; the use of the new tariff schedule in Egypt from 1931 to 1938, in India from 1931 to 1937, and in Thailand from 1927 to 1938. The coefficients associated with these dummy variables indicate that for Chile and India the incremental revenues generated were large compared to those for Egypt and Thailand. Changes in accounting practices of the railroads are measured by the dummy *NET* in Ceylon, India, and Nigeria and the dummy *RAIL* in Jamaica.

Examining the results of the expenditure equation, the average short-run revenue elasticity is .62 and the average elasticity associated with lagged expenditures is .38. For each of the ten countries, the resulting long-run revenue elasticity is very close to one. This value is consistent with the underlying data which indicate that each colonial government did indeed run a balanced budget in the long run. There is significant variation in the distribution between current and recurrent nominal expenditures for the ten countries. Chile, India, and Taiwan form one bloc where expenditures were financed almost entirely out of current revenues. For the other seven countries, recurrent expenditures were much more important.

III. Concluding Comments

A dynamic simulation of the model for each country for the full estimation period was performed to judge its performance in explaining the observed development pattern and to check the reliability of our estimates. An examination of the simulation plots for each country revealed no systematic divergences between the calculated and observed values of each variable. For example, the dynamic simulations had an average standard error of .1052 for real exports. This average compares quite well with the average standard errors of .1078 in the estimated supply equations and of .1035 in the estimated demand equations. As we have found, the historic pattern of export prices was very volatile due to relatively price inelastic and quite shiftable colonial supply and demand schedules. Thus a favorable test of the model is that the dynamic simulation of the export price for each country closely corresponded with the actual fluctuations of the export price, with .1458 as the average standard error. The simulation also indicated a check of the method of calculating accumulated real government expenditures, for which the standard error was .0436 and the average absolute residual for the last year was .0362. These dynamic simulations provide the final empirical support for the specification of the theoretical model and for the accuracy of the structural estimates.

We have found that there was a similar process of colonial development, as specified by our structural model. This model has identified common internal and ex-

ternal forces which transformed colonial countries in a similar pattern during the period of our study. A major finding is the role of government expenditures in promoting the growth of colonial export supply. We have also verified the role of developed countries' demand variables and trade policies, and we have measured how these external forces affected colonial export demand. Although this study described a similar development process, the results indicate significant differences in the coefficient estimates[24] which explain this common process. In fact, it is these very differences which explain why the same model works well in explaining the wide range of observed export and government growth rates. Thus, to understand the similar, but not identical, historical process of colonial development requires analyzing the common circular structure of their trade and government sectors as well as recognizing the diversity within that structure.

Appendix

Econometric Method

The estimation method used in this paper considers a set of K simultaneous equations

$$\text{(A.1)} \qquad Y\Gamma + Y_{-1}A + XB = U$$

where there are K endogenous variables, Y; K lagged endogenous variables, Y_{-1}; and M exogenous variables, X. For T observations, then, Y and Y_{-1} are TxK matrices, and X is a TxM matrix. Γ, A, and B are matrices of coefficients to be estimated with dimensions KxK, KxK, and MxK, respectively.

Assume the error matrix U follows a first-order autoregressive pattern

$$\text{(A.2)} \qquad U = U_{-1}R + E$$

where U_{-1} is the matrix of U lagged, and U, U_{-1}, and E are TxK matrices. Denoting e_t' as the column components of E, the following assumptions are made:

$E(e_t) = 0 \quad t = 1, 2, \ldots, T$
$E(e_t)(e_t') = \Sigma \quad t = 1, 2, \ldots, T$, Σ positive definite
$E(e_t)(e_\tau') = 0 \quad t, \tau = 1, 2, \ldots, T$, $t \neq \tau$
$plim\ T^{-1}XE = plim\ T^{-1}X_{-1}E = 0$
$plim\ T^{-1}Q'Q$ exists as a fixed, nonsingular matrix where $Q = (X, X_{-1})$
R is a diagonal matrix with elements $|r_{ii}| < 1$, $i = 1, 2, \ldots, K$
$(\Gamma + A)$ has an inverse.

The equations of the model are identified.

Without any loss of generality, let the equation to be estimated be the equation for the first endogenous variable y_1

$$\text{(A.3)} \qquad y_1 = Y_1\gamma_1 + Y_{1_{-1}}\alpha_1 + X_1\beta_1 + u_1$$

where y_1 is a column vector of T observations; Y_1 is a Txk_1 matrix of observations on k_1 other included endogenous variables; $Y_{1_{-1}}$ is a Txj_1 matrix of observations on j_1 included lagged endogenous variables; X_1 is a Txm_1 matrix on m_1 included exogenous variables; and β_1, α_1, and γ_1 are vectors of coefficients to be estimated. The T-component column vectors of error terms, u_1, $u_{1_{-1}}$, and e_1, satisfy

$$\text{(A.4)} \qquad u_1 = \rho_1 u_{1_{-1}} + e_1$$

where ρ_1 is the first diagonal element of R. As the only lagged endogenous variables appearing in this paper's theoretical model are lagged left-hand variables, equation (A.3) can be simplified to

$$\text{(A.5)} \qquad y_1 = Y_1\gamma_1 + y_{1_{-1}}\alpha_1 + X_1\beta_1 + u_1$$

From (A.4) and (A.5), the equation to be estimated is

$$\text{(A.6)} \qquad \begin{aligned} y_1 - \rho_1 y_{1_{-1}} = {} & (Y_1 - \rho_1 Y_{1_{-1}})\gamma_1 \\ & + (y_{1_{-1}} - \rho_1 y_{1_{-2}})\alpha_1 \\ & + (X_1 - \rho_1 X_{1_{-1}})\beta_1 \\ & + e_1 \end{aligned}$$

Outline of the Estimation Procedure

The equation (A.6) can be consistently estimated using the following limited information method.

(i) *Instrumental Adjustment.* Using a set of instrumental variables that include at

[24] This conclusion was statistically verified using covariance analysis.

least X_1 and $X_{1_{-1}}$ and are asymptotically uncorrelated with e_1, instrumentally adjust Y_1, $Y_{1_{-1}}$, $y_{1_{-1}}$, and $y_{1_{-2}}$ and denote their predicted values as, respectively, $\hat{Y}_1$, $\hat{Y}_{1_{-1}}$, $\hat{y}_{1_{-1}}$, and $\hat{y}_{1_{-2}}$.

(ii) *Initial Consistent Estimation.* Estimate equation (A.5) by ordinary least squares using $\hat{Y}_1$ instead of Y_1 and $\hat{y}_{1_{-1}}$ instead of $y_{1_{-1}}$. Denote the estimated coefficients of this regression as $\hat{\gamma}_1$, $\hat{\alpha}_1$, and $\hat{\beta}_1$. By the properties of instrumental variables, these estimates are consistent. A consistent estimate of the residuals u_1 is given by

$$\text{(A.7)} \quad \hat{u}_1 = y_1 - Y_1\hat{\gamma}_1 - y_{1_{-1}}\hat{\alpha}_1 - X_1\hat{\beta}_1$$

A consistent estimator of ρ_1 is given by

$$\text{(A.8)} \quad \hat{\rho}_1 = (\hat{u}_{1_{-1}}{}'\hat{u}_1)/(\hat{u}_{1_{-1}}{}'\hat{u}_{1_{-1}})$$

Thus $\hat{\rho}_1$ is the estimated coefficient obtained by regressing $\hat{u}_1$ on $\hat{u}_{1_{-1}}$.

(iii) *Iterative Procedure.* For any consistent estimator $\hat{\rho}_1$ of ρ_1, estimate (A.6) by ordinary least squares using the instrumentally adjusted endogenous variables, $\hat{y}_{1_{-1}}$, $\hat{Y}_1$, $\hat{Y}_{1_{-1}}$, and $\hat{y}_{1_{-2}}$ instead of $y_{1_{-1}}$ Y_1, $Y_{1_{-1}}$, and $y_{1_{-2}}$, respectively. The estimated coefficients of this regression are denoted by $\hat{\gamma}_1$, $\hat{\alpha}_1$, and $\hat{\beta}_1$, and these estimates are consistent. Use these estimates to calculate successively $\hat{u}_1$ and $\hat{\rho}_1$, using equations (A.7) and (A.8). Starting with the initial consistent estimates of (ii), continue this iterate procedure until the estimates converge.

(iv) *Equation Statistics.* A consistent estimator of e_1 is

$$\text{(A.9)} \quad \hat{e}_1 = \hat{u}_1 - \hat{\rho}_1\hat{u}_{1_{-1}}$$

Calculate the standard error of the equation using

$$\text{(A.10)} \quad \hat{S}_1 = \sqrt{(\hat{e}_1'\hat{e}_1)/T}$$

The estimated variance-covariance matrix of the estimated coefficients is

$$\text{(A.11)} \quad V(\hat{\delta}_1) = \hat{S}_1^2(\hat{Z}_1'\hat{Z}_1)^{-1}$$

where: $\hat{\delta}_1 = [\hat{\gamma}_1, \hat{\alpha}_1, \hat{\beta}_1]$ and $\hat{Z}_1 = [\hat{Y}_1 - \hat{\rho}_1\hat{Y}_{1_{-1}}, \hat{y}_{1_{-1}} - \hat{\rho}_1\hat{Y}_{1_{-2}}, X_1 - \hat{\rho}_1 X_{1_{-1}}]$

Characteristics of the Estimation Procedure

Consistency of the procedure relies on three facts. First, all the included predetermined variables appearing in equation (A.6), namely X_1 and $X_{1_{-1}}$, appear in the list of instruments used in step (i). Second, the equation is assumed to be identified. Third, the set of instrumental variables used are asymptotically uncorrelated with e_1.

In this estimation procedure, the lagged endogenous variables appearing in (A.6), $Y_{1_{-1}}$, $y_{1_{-1}}$, and $y_{1_{-2}}$, are treated as endogenous rather than exogenous. Recalling the discussion by Franklin Fisher, the endogeneity of these lagged variables arises from the fact that they are correlated with the error term, e_1. This argument is particularly applicable here because the lagged left-hand variable appears due to both a distributed lag process and an autoregressive process. If $Y_{1_{-1}}$, $y_{1_{-1}}$, and $y_{1_{-2}}$ are endogenous, then the estimation methods proposed by Ray Fair and Phoebus Dhrymes will yield inconsistent estimators since their methods treat lagged endogenous variables as exogenous.

The instrumental adjustment was done using the structurally ordered instrumental variable method described by Fisher and by Bridger Mitchell and Fisher. In the econometric model, structural ordering emphasizes lagged exogenous variables rather than current exogenous variables in the instrumental adjustment of lagged endogenous variables, while still satisfying the consistency requirement of using X_1 and $X_{1_{-1}}$ as instruments. First regress Y_1, $Y_{1_{-1}}$, $y_{1_{-1}}$, and $y_{1_{-2}}$ on instrument lists of different exogenous variables. Let $\tilde{Y}_1$, $\tilde{Y}_{1_{-1}}$, $\tilde{y}_{1_{-1}}$, and $\tilde{y}_{1_{-2}}$ be fitted values of these regressions. For consistent estimators, then instrumentally adjust Y_1, $Y_{1_{-1}}$, $y_{1_{-1}}$, and $y_{1_{-2}}$ using $\tilde{Y}_1$, $\tilde{Y}_{1_{-1}}$, $\tilde{y}_{1_{-1}}$, $\tilde{y}_{1_{-2}}$, X_1 and $X_{1_{-1}}$ to obtain the predicted values $\hat{Y}_1$, $\hat{Y}_{1_{-1}}$, $\hat{y}_{1_{-1}}$. and $\hat{y}_{1_{-2}}$ of the first step of the estimation procedure.

Definitions of Developed Country Variables and Dummy Variables

Developed Country Variables

- CAR = Motor vehicle factory sales, United States
- P_{JP} = GNP price deflator with 1934–36 = 100, Japan
- P_{UK} = National income price deflator with 1913–14 = 100, United Kingdom

P_{US} = GNP price deflator with 1929 = 100, United States

P^m_{US} = Fisher import price index with 1913 = 100, United States

P^{mm}_{US} = Import price index of crude materials with 1913 = 100, United States

Q_{UK} = Index of industrial production excluding building, United Kingdom

Q_{US} = Index of manufacturing output with 1929 = 100, United States

Y^R_{JP} = Real GNP in millions of 1934–36 yen, Japan

Y^R_{UK} = Real net national income in millions of 1913–14 pounds, United Kingdom

Y^R_{US} = Real GNP in millions of 1929 dollars, United States

Dummy Variables

FIXED = Thailand, tariffs fixed by the Bowring Treaty until 1926, and thereafter increasing tariffs, 1927–37

INCOME = Chile, income tax on copper producers, 1926–38

INFRA = Nigeria, completion of infrastructure projects—railroad to Northern Nigeria and port of Lagos—1917–38

LIMIT = Ceylon, international restriction scheme on rubber exports, 1935–38

NET = Change from gross to net railway revenues in Ceylon, 1929–38; India, 1906–36; Nigeria, 1927–38

QUOTA = Cuba, *U.S.* import sugar quotas and tariffs, 1930–37; Philippines, *U.S.* import sugar quotas, 1935–38

RAIL = Jamaica, government takeover railroads, 1900–38

RESTR = *U.S.* import tariffs and restrictions in Chile, 1932–38; in Jamaica, 1932–38

TARIFF = New tariff schedules in Egypt, 1931–38; India, 1931–37

WORKS = Ceylon, expenditure includes public works expenses, 1916–24

WWI = Effects of World War I in Ceylon, 1915–18; Egypt, 1915–18; Nigeria, 1915–21; Taiwan, 1915–19

Appendix Table 1—Export Supply Equation Estimates[a]

Country	Constant	$ln\ Px$	$ln\ Pm$	ARGE	$ln\ X^R_{-1}$	Dummy Variables	R^2	ρ	D.W.	S.E.
Ceylon	−4.15	.303	−.144	.349	.842	−.191 *LIMIT*	.987	.0	2.02	.069
	(−3.0)	(3.6)	(−1.5)	(1.9)	(6.1)	(−2.9)				
Chile	7.35	.465	−.443	.572		−.525 *RESTR*	.789	.430	1.97	.195
	(4.0)	(1.8)	(−1.8)	(6.5)		(−1.9)				
Cuba	1.76	.222	−.342	.489	.373	−.488 *QUOTA*	.854	.0	2.25	.123
	(0.8)	(1.0)	(−1.2)	(3.3)	(1.8)	(−3.0)				
Egypt	6.02	.267	−.454	.316		.146 *WWI*	.542	.0	1.88	.111
	(6.8)	(1.4)	(−2.4)	(4.9)		(1.1)				
India	14.89	.422	−.286	.267			.754	.597	1.60	.077
	(4.6)	(2.2)	(−1.9)	(2.1)						
Jamaica	4.10	.536	−.665	.606		−.226 *RESTR*	.709	.258	1.97	.148
	(2.5)	(1.6)	(−2.4)	(6.4)		(−1.5)				
Nigeria	1.76	.142	−.233	.361	.495	.268*INFRA* − .081*WWI*	.982	.0	2.04	.069
	(2.3)	(1.3)	(−2.2)	(3.1)	(2.8)	(2.8) (−2.0)				
Philippines	−.20	.268	−.220	.921		−.173 *QUOTA*	.965	.257	1.80	.110
	(−.2)	(1.2)	(−0.9)	(20.7)		(−2.0)				
Taiwan	−2.01	.363	−.482	.453	.626	.108 *WWI*	.982	−.243	2.11	.086
	(−.7)	(1.6)	(−2.1)	(1.4)	(3.1)	(2.7)				
Thailand	2.52	.158	−.209	.361			.918	.0	2.12	.091
	(5.2)	(0.7)	(−1.0)	(16.7)						

Appendix Table 2—Demand Equation Estimates[a]

Country	Constant	$ln\ Px$	Developed Country's Income	Developed Country's Prices	$ln\ X^R_{-1}$	Dummy Variable	R^2	ρ	$D.W.$	$S.E.$
Ceylon	9.89 (4.7)	−.441 (−5.3)	.787 $ln\ Y^R_{UK}$ (2.8)	.255 $ln\ CAR$ (15.)		−.210$LIMIT$ (−2.7)	.985	.614	1.59	.075
Chile	7.51 (2.7)	−1.011 (−4.3)	.532 $ln\ Q_{US}$ (3.8)	.759 $ln\ P^{nm}_{US}$ (3.9)	.333 (2.2)	−.564$RESTR$ (2.7)	.828	−.133	2.06	.176
Cuba	10.98 (10.)	−.784 (−4.1)	.647 $ln\ Q_{US}$ (4.8)	1.154 $ln\ P^{m}_{US}$ (4.0)		−.350$QUOTA$ (−3.3)	.883	.351	1.79	.112
Egypt	3.52 (3.8)	−.228 (−3.5)	.882 $ln\ Y^R_{UK}$ (7.2)	.169 $ln\ P_{UK}$ (2.2)		−.163 WWI (−3.2)	.687	.0	2.12	.092
India	18.05 (15.)	−.214 (−1.3)	.604 $ln\ Q_{UK}$ (3.2)	.151 $ln\ P_{UK}$ (1.0)			.716	.752	1.93	.083
Jamaica	10.04 (14.)	−.451 (−2.3)	1.031 $ln\ Q_{UK}$ (6.5)	.236 $ln\ P_{UK}$ (1.3)		−.121$RESTR$ (−1.4)	.790	.136	1.94	.126
Nigeria	−.60 (−.5)	−.189 (−2.5)	.617 $ln\ Y^R_{UK}$ (2.3)	.366 $ln\ P_{UK}$ (2.8)	.735 (7.8)	−.097WWI (−2.3)	.979	.0	2.09	.076
Philippines	3.14 (1.6)	−.498 (−2.4)	.424 $ln\ Y^R_{US}$ (1.8)	.601 $ln\ P_{US}$ (2.3)	.592 (3.3)	−.056$QUOTA$ (−1.0)	.976	−.216	2.04	.090
Taiwan	2.21 (0.6)	−1.095 (−2.9)	1.650 $ln\ Y^R_{JP}$ (4.8)	.287 $ln\ P_{JP}$ (1.0)		.171WWI (1.8)	.966	.802	2.07	.118
Thailand	−3.66 (−2.0)	−.212 (−2.0)	1.093 $ln\ Y^R_{UK}$ (4.9)	.736 $ln\ P_{UK}$ (8.2)			.899	.251	1.87	.087

Appendix Table 3—Import Demand Equation Estimates[a]

Country	$ln\ X^R$	$ln\ Pm$	$ln\ Px$	Dummy Variable	R^2	ρ	$D.W.$	$S.E.$
Ceylon	.983 (135.)	−.366 (−2.5)	.334 (2.6)	−.204WWI (−3.1)	.946	.903	2.12	.080
Chile	.979 (263.)	−.411 (−1.8)	.700 (3.6)		.779	.302	1.93	.211
Cuba	.985 (183.)	−.171 (−0.5)	.363 (1.5)		.801	.745	1.84	.140
Egypt	.986 (103.)	−.403 (−1.1)	.277 (0.8)	−.177WWI (−0.8)	.837	.727	2.30	.174
India	.979 (293.)	−.450 (−1.6)	.405 (1.2)		.815	.666	1.80	.124
Jamaica	1.018 (102.)	−.581 (−1.8)	.855 (2.3)		.869	.850	2.34	.152
Nigeria	.994 (111.)	−.941 (−5.4)	.750 (3.6)	−.259WWI (−2.6)	.809	.720	2.02	.119
Philippines	.986 (377.)	−.024 (−0.1)	.256 (1.3)		.941	.421	1.86	.129
Taiwan	.975 (236.)	−.288 (−1.1)	.182 (0.6)		.967	.504	1.52	.114
Thailand	.928 (37.)	−.548 (−1.7)	.416 (1.5)		.903	.822	2.12	.121

Appendix Table 4—Government Revenue Equation Estimates[a]

Country	Constant	$ln\ X^R$	$ln\ M$	Dummy Variables		R^2	ρ	D.W.	S.E.
Ceylon	−0.26	.754	.192	−.037*WORKS*	−.039*NET*	.973	.510	2.01	.088
	(−0.2)	(6.1)	(1.4)	(−.6)	−.5)				
Chile	−4.77	1.203		1.310*INCOME*		.964	.412	2.00	.188
	(−1.8)	(8.7)		(11.)					
Cuba	−3.06	.913	.191			.840	.612	1.85	.156
	(−0.7)	(3.3)	(1.8)						
Egypt	−2.80	.636	.611	.032*TARIFF*		.945	.650	2.04	.116
	(−1.0)	(2.7)	(7.5)	(1.3)					
India	−1.83	.278	.813	.564*TARIFF*	−.332*NET*	.920	.443	1.85	.091
	(−0.4)	(1.2)	(8.0)	(7.8)	(−.3.7)				
Jamaica	−0.57	.353	.627	.152*RAIL*		.976	.250	2.01	.068
	(−0.6)	(4.4)	(16.)	(3.8)					
Nigeria	−6.85	.889	.487	−.262*NET*		.979	.441	1.80	.091
	(−5.7)	(7.9)	(6.5)	(−3.1)					
Philippines	−0.43	.388	.603			.964	.313	1.94	.101
	(−0.5)	(5.3)	(7.4)						
Taiwan	5.01	.314	.357			.940	.411	2.20	.117
	(4.2)	(1.8)	(2.2)						
Thailand	1.70	.230	.325	.020*FIXED*		.941	.672	1.67	.062
	(2.0)	(1.2)	(4.1)	(1.9)					

Appendix Table 5—Government Expenditures Equation Estimates[a]

Country	$ln\ R$	$ln\ G_{-1}$	*WORKS*	R^2	ρ	D.W.	S.E.
Ceylon	.505	.496	.029	.981	−.203	2.01	.076
	(3.7)	(3.6)	(1.1)				
Chile	.951	.047		.982	.428	1.70	.139
	(9.7)	(0.5)					
Cuba	.266	.741		.949	−.172	1.90	.100
	(4.6)	(13.)					
Egypt	.556	.443		.981	.0	2.26	.065
	(4.7)	(3.7)					
India	.847	.153		.988	.542	1.74	.037
	(8.6)	(1.6)					
Jamaica	.579	.421		.986	−.209	1.95	.055
	(4.8)	(3.5)					
Nigeria	.400	.604		.971	.296	2.13	.094
	(4.2)	(6.3)					
Philippines	.652	.348		.986	−.389	1.95	.066
	(8.5)	(4.5)					
Taiwan	.851	.146		.973	.332	2.07	.084
	(7.2)	(1.2)					
Thailand	.626	.367		.960	.526	2.16	.058
	(4.4)	(2.5)					

[a] t-values in parentheses.

REFERENCES

A. E. Crouchley, *The Economic Development of Modern Egypt*, London 1938.

P. Deane and W. A. Cole, *British Economic Growth, 1688–1959*, Cambridge 1967.

P. J. Dhrymes, "Full Information Estimation of Dynamic Simultaneous Equations Models with Autoregressive Errors," mimeo., Mar. 1971.

G. Eisner, *Jamaica, 1830–1930: A Study in Economic Growth*, Manchester 1961.

R. C. Fair, "The Estimation of Simultaneous Equation Models with Lagged Endogenous Variables and First Order Serially Correlated Errors," *Econometrica*, May 1970, *37*, 507–16.

F. M. Fisher, "Dynamic Structure and Estimation in Economy-Wide Econometric Models," in J. Dusenberry et. al., eds., *The Brookings Quarterly Econometric Model of the United States*, Chicago 1965, 588–635.

G. Helleiner, *Peasant Agriculture, Government and Economic Growth in Nigeria*, Homewood 1966.

S. Ho, *Statistical Appendix to Taiwan*, unpublished.

S. Hymer and S. Resnick, "Interaction Between the Government and the Private Sector: An Analysis of Government Expenditure Policy and the Reflection Ratio," in I. G. Stewart, ed., *Economic Development and Structural Change*, Edinburgh 1969, 155–80.

J. Ingram, *Economic Change in Thailand Since 1850*, Stanford 1955.

C. Issawi, *Egypt: An Economic and Social Analysis*, Oxford 1947.

R. E. Lipsey, *Price and Quantity Trends in the Foreign Trade of the United States*, Princeton 1963.

M. Mamalakis and C. Reynolds, *Essays on the Chilean Economy*, Homewood 1965.

D. Mead, *Growth and Structural Change in the Egyptian Economy*, Homewood 1967.

B. M. Mitchell and F. M. Fisher, "The Choice of Instrumental Variables in the Estimation of Economy-Wide Econometric Models: Some Further Thoughts," *Int. Econ. Rev.*, June 1970, *11*, 226–33.

B. R. Mitchell, *Abstract of British Historical Statistics*, Cambridge 1962.

S. Resnick, "The Decline of Rural Industry Under Export Expansion: A Comparison among Burma, Philippines and Thailand," *J. Econ. Hist.*, Mar. 1970, *30*, 51–73.

———, *Statistical Appendix to the Philippines*, unpublished.

D. Snodgrass, *Ceylon: An Export Economy in Transition*, Homewood 1966.

Chile, Officina Central de Estadistica, *Sinopsis Estadistica de la Republic de Chile*, Santiago 1893–1939.

Cuba, Secretaria de Hacienda, Seccion de Estadistica, *Commercio Exterior*, Habana 1903–38.

———, *Annuario Estadistica de la Republic de Cuba, 1*, Habana 1915.

The Cuban Economic Research Project, *A Study on Cuba*, Coral Gables 1965.

Great Britain, Board of Trade, *Statistical Abstract for the British Empire* (title varies), *33–68*, London 1887–1939.

———, *Statistical Abstract for the Principal and Other Foreign Countries, 15–39*, London 1889–1914.

League of Nations, *International Trade Statistics*, Geneva 1932–39.

———, *Memorandum on Balance of Payments and Foreign Trade Balances*, Geneva 1924–27.

———, *Memorandum on International Trade and Balance of Payments*, Geneva 1928–31.

Siam, Ministry of Finance, Department of General Statistics, *Statistical Yearbook of the Kingdom of Siam, 1–21*, Bangkok 1914–44.

The Statesman's Yearbook, London 1902–38.

U.S. Bureau of the Census, *Historical Statistics of the U.S., Colonial Times to 1957*, Washington 1960.

Heavy and Light Industry in Economic Development

By MORRIS TEUBAL*

Accepted doctrine regarding the choice between light (labor-intensive) industry and heavy (capital-intensive) industry in the early stages of industrialization reflects in part the conclusions of the neoclassical Factor Proportion Theorem and the Rybczynski Theorem.[1] According to these, in a small economy (i.e., one which has no influence on world prices) under free trade, light industry must necessarily come into existence before heavy industry.

This prescribed optimum industrialization sequence is not necessarily in accordance with the experience of European industrialization; Alexander Gerschenkron, in particular, has pointed out that there is a variety of industrialization patterns including instances of initial emphasis on heavy industry. Our purpose here is to present a model in which the optimum strategy may be initially based on either light or heavy industry, depending on the factor endowment, technology, and factor accumulation patterns of the country involved. It is hoped in this way to contribute to a better understanding of the factors which should be taken into account when devising desirable industrialization strategies for today's developing countries.[2]

The analysis is carried out in terms of a three-sector neoclassical model of an economy which faces fixed, exogenously determined prices for all goods. Except for the explicit introduction of an agricultural sector, I follow the standard Hecksher-Ohlin model in assuming free trade with no transport costs, two factors of production (labor and capital), constant returns (except in the primary sector), a given technology, and no externalities. Thus, it resembles the model presented in my previous paper except that the relative price of the primary good is determined exogenously in world markets (small economy assumption) instead of being determined by the quantity of exports (medium-sized economy assumption).[3]

The conclusions about dynamic comparative advantage that emerge from the model are quite clear-cut despite the fact that an accumulation mechanism is not explicitly considered. In particular, heavy industry may be the optimum starting point for industrialization, as pointed out in my earlier paper. This possibility is greater the higher (lower) the initial capital stock (labor force), the higher (lower) the rate of capital accumulation (population growth) and, everything else equal, the higher the returns to scale in the primary sector (within the class of decreasing returns to scale production functions).[4]

* Lecturer, department of economics, Eliezer Kaplan School of Economics and Social Sciences, Hebrew University.

[1] See, for example, R. W. Jones and T. M. Rybczynski.

[2] In what follows I draw on an earlier article published in 1971.

[3] This guarantees the independence of the allocation decision from the saving or accumulation decision. In the previous model this was not possible; there, an increase in investment is in part financed by an increase in exports of primary goods, which reduces their relative price and induces resource reallocation.

[4] These results are an extension of those obtained previously. In that paper a constant labor force was assumed although an explicit, utility maximizing mechanism of accumulation was postulated. Some proofs relevant to both papers are given in the Appendix.

I

Technology

The technology of the three sectors is described in my 1971 paper. There are two augmentable factors of production—capital and labor—and the production functions of the industrial sectors (B and C for light and heavy industry, respectively) are homogeneous of degree one, while that of the primary sector A, is of the decreasing returns to scale variety (because land is nonaugmentable and/or of declining fertility). Let K_i, L_i, k_i, y_i be the capital input, labor input, capital-labor ratio, and output of sector i, respectively. Then

$$(1) \qquad y_a = F^a(K_a, L_a) \equiv F^a(k_aL_a, L_a)$$

$$(2) \qquad \lambda F^a(K_a, L_a) > F^a(\lambda k_aL_a, \lambda_a) \qquad \text{for } \lambda > 1$$

$$(3) \qquad y_i = F^i(K_i, L_i) = L_if_i(k_i)$$

It is further assumed for the primary sector that (i) any increase in scale (that is, and increase in L_a for constant k_a) reduces the marginal product of *both* factors; (ii) the marginal product of one factor, say labor, increases to infinity as the scale approaches zero; and (iii) factors are complementary. Let $F_J^a(J=K, L)$ be the marginal product of factor J in agriculture and F_{KK}^a, F_{LL}^a, F_{KL}^a the direct and cross second-order derivatives of the production function in this sector. Then, assumptions (i)–(iii) imply

$$(4) \qquad \frac{d}{dL_a}(F_J^a) = k_aF_{JK}^a + F_{JL}^a < 0 \qquad (J=K, L)$$

$$(5) \qquad \lim_{\substack{L_a\to 0\\ k_a>0}} F_L^a(k_aL_a, L_a) \to \infty$$

$$(6) \qquad F_{KL}^a = F_{LK}^a > 0$$

Equation (4) defines a subclass of decreasing returns to scale production functions, i.e., it is sufficient for condition (2) to hold. Equation (5) guarantees that for a positive world price of agricultural goods, agricultural output is necessarily positive.[5] Finally assume $k_c > k_b > k_a$.

Maximization

Let Y, K, L be national income at world prices in terms of heavy-industry goods, total capital, and total labor, respectively. At any moment of time, the objective is an allocation of (K, L) which maximizes Y, i.e.,

$$\max_{(K_i,L_i)} Y = p_aF^a(K_a, L_a) + [qF^b(K_b, L_b) + F^c(K_c, L_c)]$$

subject to $K_b + K_c = K - K_a$ and $L_b + L_c = L - L_a$, where p_a, q are the world prices of agricultural and light-industry goods in terms of heavy-industry goods.

The first-order conditions for a maximum under the assumption that all three goods are being produced in positive quantities, i.e., that the economy is in the nonspecialization phase, can be grouped as follows:

i. *Marginal productivity conditions of the industrial sectors*

$$(7) \qquad q\frac{\partial F^b}{\partial K_b}(k_b) = \frac{\partial F^c}{\partial K_c}(k_c) = r$$

$$(8) \qquad q\frac{\partial F^b}{\partial L_b}(k_b) = \frac{\partial F^c}{\partial L_c}(k_c) = w$$

ii. *Marginal productivity conditions of the agricultural sector*

$$(9) \qquad p_aF_K^a(K_a, L_a) = r$$

$$(10) \qquad p_aF_L^a(K_a, L_a) = w$$

iii. *Resource constraints*

$$(11) \qquad k_bL_b + k_cL_c = K - K_a$$

$$(12) \qquad L_b + L_c = L - L_a$$

[5] Assumption (ii) or equation (5) is sufficient but not necessary for the purpose of the paper. The fact that it guarantees a positive output of agricultural goods simplifies the exposition without affecting qualitatively the results on the optimum strategy in the early stages of industrialization. See fn. 12 below.

In addition, we have the *nonnegative restriction* on sectoral capital and labor allocations (K_i, $L_i \geq 0$).

II

General Equilibrium under Nonspecialization[6]

To understand the structure of factor allocation under nonspecialization, it is useful to distinguish between the capital-labor ratio of the economy, k, and the aggregate capital-labor ratio of the industrial sectors, k'. Let the difference between the capital (labor) endowment and the quantity demanded by the primary sector under nonspecialization be $S_K(S_L)$ (i.e., "surplus" capital (labor)). Then

$$k' \equiv \frac{S_K}{S_L} = \frac{kL - \hat{K}_a}{L - \hat{L}_a} \tag{13}$$

In this expression, $\hat{K}_a$ and $\hat{L}_a$ will be constants. This is because together with $\hat{k}_i(i=b, c)$, and $\hat{r}$, $\hat{w}$ they are determined by marginal productivity conditions (7)–(8), (9)–(10), and they are thus independent of factor endowments and depend only on prices and technology which are assumed given. Thus we put $\hat{K}_a$ $K_a(p_a, q)$ and $\hat{L}_a = L_a(p_a, q)$.[7]

The allocation of surplus factors to the industrial sectors will now be determined as is usual in two sector models, with k' replacing k. Let us write $k_i(q)$ for $\hat{k}_i(i=b, c)$ and $k'(k, L)$ for k'. Then resource constraints (11)–(12) imply the following condition for nonspecialization, i.e., for $\hat{L}_b$, $\hat{L}_c > 0$ (provided S_K and S_L are both positive):

$$k_c(q) > k'(k, L) > k_b(q) \tag{14}$$

Condition (14) states that a necessary and sufficient condition for a positive output of both industrial goods is for the aggregate industrial capital-labor ratio to lie between the capital-labor ratios of the heavy-industry and light-industry sector. The $S_K = 0$ and $S_L = 0$ schedules are drawn in Figure 1; they intersect at $E[L_a(p_a, q), k_a(p_a, q) \equiv K_a(p_a, q)/L_a(p_a, q)]$. The area of (k, L) where S_K, $S_L > 0$ is the area to the right of $S_K = 0$ which lies above $S_L = 0$. The $k_c(q) = k'(k, L)$ schedule (which corresponds to $\hat{L}_b = 0$) and the $k_b(q) = k(k, L)$ schedule (which corresponds to $\hat{L}_c = 0$) are also drawn. Both intersect at E and are upward sloping since, from (13)

$$\frac{\partial k'}{\partial k} = \frac{L}{L - L_a(p_a, q)} > 0 \qquad (\text{for } S_L > 0)$$

$$\frac{\partial k'}{\partial L} = -\frac{L_a(p_a, q)}{[L - L_a(p_a, q)]^2}[k - k_a(p_a, q)] < 0^{8}$$

Condition (14) holds in the shaded area which lies between these two schedules, and this will therefore be the nonspecialization region. The next step is to determine the regions in (k, L) where only one industrial good should be produced in addition to the primary good—industrial specialization—and where no industrial good should be produced—agricultural specialization. We start with industrial specialization in, say, light industry. A necessary condition for this pattern to be optimum is for the marginal productivity conditions (holding with equality) of the light-industry and agricultural sectors to be consistent with the resource constraints for $L_b > 0$, $L_c = 0$ (we already know that L_a must always be positive). This is *feasibility*, a necessary condition for optimality. *Optimality* of light-industry specialization in addition requires the shadow prices of

[6] The circumflex identifies the variables of this phase of specialization.

[7] Thus the output of the primary sector and the production techniques of all sectors are constant under nonspecialization.

[8] Thus an increase in the size of the economy (i.e., a proportional increase in both capital and labor) will increase the share of light industry in total industrial output. This effect is similar to the effect of "growth acceleration" in my earlier article.

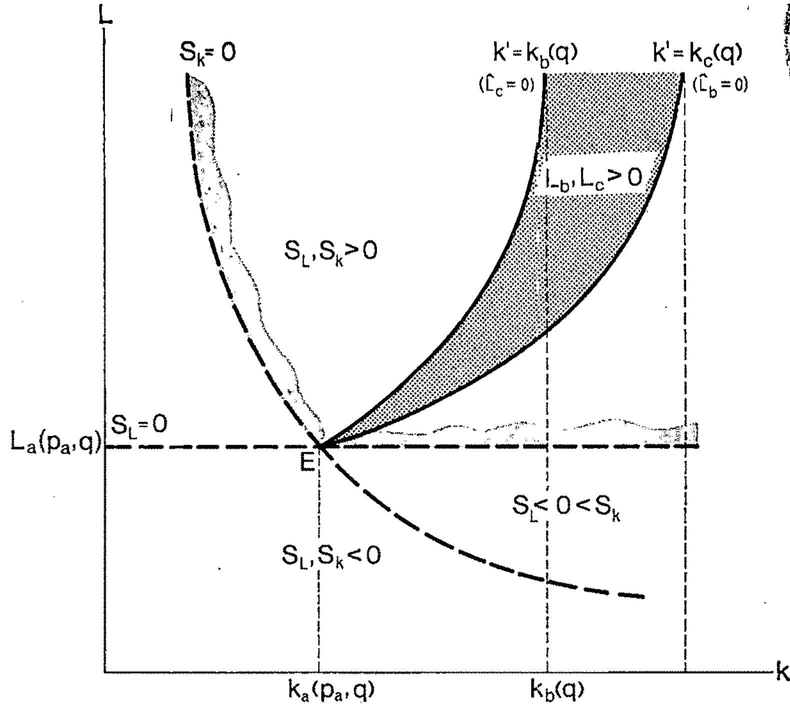


FIGURE 1

capital and labor in this phase of specialization to be equal to or greater than the marginal product of capital and labor in the nonproducing (heavy-industry) sector (at least one of these conditions must hold with strict inequality).

Feasibility of Light-Industry Specialization[9]

From marginal conditions (7), (8), (9), and (10) we obtain an implicit relationship between the sectoral capital-labor ratios in this phase, $\bar{k}_i$ and primary-sector labor L_a:

$$\bar{k}_i = \bar{k}_i(L_a) \qquad (i = a, b) \tag{15}$$

It follows from differentiation of the above marginal conditions that $\bar{k}_i'(L_i) < 0$, i.e., that the marginal conditions for light industry and agriculture imply an increase in primary-sector labor to be accompanied by a reduction in the capital-labor ratios of both sectors.[10] This is shown in Figure 2A.

To obtain the actual labor allocations we introduce (15) into resource constraints (11) and (12) for $L_c = 0$:

$$\frac{L_a}{L} = \frac{\bar{k}_b(L_a) - k}{\bar{k}_b(L_a) - \bar{k}_a(L_a)} \tag{16}$$

This equation determines the actual labor allocation to the primary sector as a function of (k, L). We will denote it by $\bar{L}_a(k, L)$. Introducing this value into (15), we obtain the sectoral capital-labor ratios.

Feasibility of light-industry specializa-

[9] The variables of this phase of specialization are barred.

[10] The actual expression for $\bar{k}_a(L_a)$ is:

$$\bar{k}_a(L_a) = -\frac{\bar{k}_a}{L} \cdot \frac{\frac{d}{dL_a}(F_k^a) + \frac{d}{dL_a}(F_L^a)}{k_b F_{KK}^a + F_{KL}^a} < 0$$

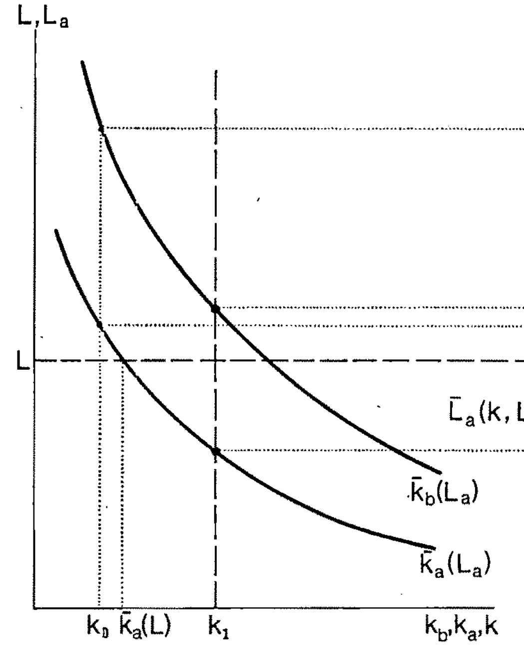


FIGURE 2A

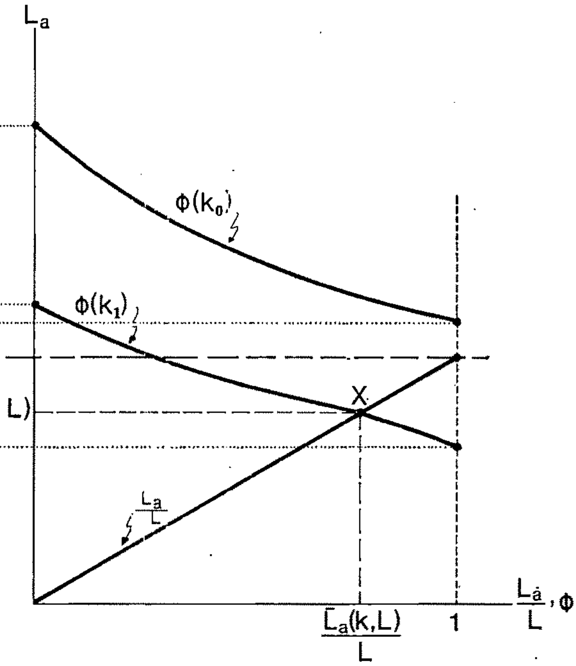


FIGURE 2B

tion means

$$(17) \qquad \bar{L}_a(k, L) < L.$$

We will show that a necessary and sufficient condition is $k > \bar{k}_a(L)$. Suppose that $k = k_0 < \bar{k}_a(L)$. It is clear in this case from Figure 2A that L_a/L (the left-hand side of (16)) less than unity implies that the right-hand side of (16)—which we will denote by ϕ—is greater than unity. Thus, light-industry specialization is not feasible when $k < \bar{k}_a(L)$. On the other hand, when $k > \bar{k}_a(L)$ there will necessarily be a solution to (16) which satisfies (17). To show this we represent both L_a/L and ϕ as functions of L_a in Figure 2B. At the level of L_a which makes $\bar{k}_b(L_a)$ equal to the capital-labor ratio of the economy, say k_1, we have $\phi = 0$. On the other hand, when $k_1 = \bar{k}_a(L_a)$ we have $\phi = 1$. At intermediate levels of L_a, ϕ will necessarily be a declining function of L_a.

The figure shows clearly that when $k = k_1 > \bar{k}_a(L)$, the ϕ and L_a/L schedules intersect at a level $\bar{L}_a(k, L) > L$ (point X). On the other hand, there is no such solution when $k = k_0 < \bar{k}_a(L)$. Thus in the (k, L) space, light-industry specialization is feasible at and only at points above and to the right of the $\bar{k}_a(L)$ schedule.

Similarly for feasibility of heavy-industry specialization.[11] We obtain a $\tilde{k}_a(L_a)$ schedule from the marginal productivity conditions of the agricultural and heavy-industry sectors. Then feasibility of this pattern of specialization implies $k > \tilde{k}_a(L)$. Will the $\bar{k}_a(L)$ and $\tilde{k}_a(L)$ schedules intersect? The answer is yes and precisely at point E. This is so because the marginal productivity conditions of *all three* sectors hold for $L_a = L$, $L_b = L_c = 0$ at both E and at the intersection of $\tilde{k}_a(L_a)$ and $\bar{k}_a(L)$. Moreover, as we have seen, this can only hold at one point. In addition, since $k_c > k_b$, it is easily shown that $\tilde{k}_a'(L)$

[11] The tilde identifies the variables of this phase.

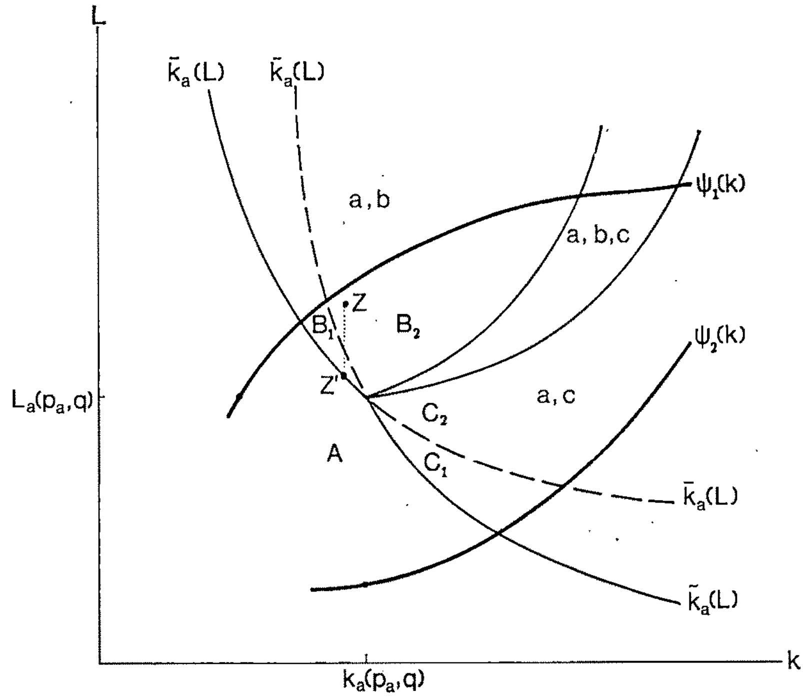


FIGURE 3

$<\bar{k}_a'(L)<0$. Hence, the $\hat{k}_a'(L)$ and $\bar{k}_a'(L)$ schedules look as in Figure 3.

Optimality of Industrial Specialization

There are five regions in the (k, L) space: A, B_1, B_2, C_1, and C_2 apart from the non-specialization region (see Figure 3). We have to find the optimum specialization patterns in each of them.

PROPOSITION 1: *Agricultural specialization is optimum if, and only if, the economy lies in A.*

PROOF:

If the economy is in A the only feasible pattern of specialization is agricultural specialization. Hence this is also the optimum pattern. In the Appendix we prove that agricultural specialization is not optimum at any point outside A (though, it is, of course, feasible).

PROPOSITION 2: *Light-industry specialization is optimum in B_1 and B_2 and heavy-industry specialization is optimum in C_1 and C_2.*

PROOF:

In B_1 the only feasible patterns are agricultural specialization and light-industry specialization. From Proposition 1, agricultural specialization is not optimum. This leaves light-industry specialization as the optimum pattern. Similarly, in C_1 the optimum is heavy-industry specialization. In addition there are a priori indications that light-industry specialization is optimum in B_2 and heavy-industry specialization in C_2. This is so because in B_2 resource constraints imply $\hat{L}_c<0$ while in C_2 they imply $\hat{L}_b<0$. A rigorous proof is given in the Appendix.

The optimum patterns of specialization

in the (L, K) space are shown in Figure 3.[12]

III

The main resource-allocation implication of the model is that there is a minimum capital-labor ratio for heavy industry and a minimum scale (or labor force) for light industry. These are $k_a(p_a, q)$ and $L_a(p_a, q)$, respectively, the capital-labor ratio of and the labor allocation to agriculture under nonspecialization.

The minimum capital-labor ratio required for heavy industry is likely to diminish when the returns to scale of the primary sector increase. Thus, when no agriculture is possible, an extreme case where there are no natural resources at all, this three-sector model with agriculture reduces to the standard two-sector model where $k_b(q) > k_a(p_a, q)$ is the minimum capital-intensity level for the development of heavy industry. Moreover, if we consider the class of agricultural production functions which are homogeneous of degree less than one, $y_a = L_a^r f_a(k_a)(0 > r > 1)$. We find, after substituting into the agricultural marginal productivity conditions, that

$$\frac{dk_a(p_a, q)}{dr} < 0, \qquad \frac{dL_a(p_a, q)}{dr} > 0$$

That is, an increase in the degree of homogeneity of the agricultural production function reduces the minimum capital-labor ratio required for the establishment of heavy industry (and increases the minimum scale required for the establishment of light industry).

The minimum scale required for light industry is part of a general differential scale effect favorable to this sector and relatively unfavorable to heavy industry. This is clearly shown within the nonspecialization sector. It is also reflected in the possible shifts from one phase to another induced by changes in scale. Consider, for example, an economy with $k_a(p_a, q) < k < k_b(q)$. At a relatively small scale, this economy should be in the phase of heavy-industry specialization. At a somewhat larger scale it shifts to nonspecialization and at a still larger scale to light-industry specialization.

PROPOSITION 3: *For the class of factor accumulation paths where population growth is associated with capital deepening, industrialization should start with light (heavy) industry if the minimum scale required for the development of light industry is attained before the minimum capital-labor ratio required for the development of heavy industry (if the minimum capital-labor ratio for the development of heavy industry is attained before the minimum scale for the development of light industry).*

The proof can be seen by observing paths $L = \psi_1(k)$ and $L = \psi_2(k)$ of Figure 3. This and the previous discussion imply:

PROPOSITION 4: *The possibility of developing heavy industry in the early stages of industrialization is greater (a) the greater the rate of capital accumulation relative to the rate of population growth; (b) the greater the initial capital stock and the lower the initial labor force; and (c) the greater (other things equal) the efficiency of the agricultural sector.*

[12] When (5) does not hold, i.e., if

$$(5') \qquad 0 < \lim_{\substack{L_a \to 0 \\ k_a > 0}} F_L^a(k_a L_a, L_a) = M < \infty$$

it may be optimal—for some factor endowments—to produce industrial goods only, provided the scale of the economy is sufficiently high. In such a case, regions in which there is no agriculture would also have to be described in Figure 3. However, the neoclassical technology will always ensure that between the region of agricultural specialization and the region with no agriculture, we will always have some region with both industry and agriculture. Hence, the same alternative transistions between agriculture and industry which are implicit in Figure 3 (and which are discussed in the next section) also hold when (5′) holds.

The commonly held view that an optimum industrialization strategy must *necessarily* start with the development of light industries is seriously put to question. This and my 1971 paper show that it is theoretically possible, both for a small economy and for a medium-sized one, for the optimum strategy to involve the development of heavy industry in the initial stages of industrialization. However, without knowledge of the orders of magnitude of the parameters of the model we cannot say at this stage whether this possibility is or is not a realistic one for today's developing countries.

The crucial condition for this variety of industrialization strategies seems to be the existence of at least one sector with non-constant returns to scale. It is interesting that this need not be an industrial sector. Moreover, the results obtained do not depend on assumptions about the nature of the agricultural production function (in addition of decreasing returns to scale) or on the assumption of perfect intersectoral mobility of capital, or on differential external effects. Finally, the optimality of developing heavy industries at the early stages of industrialization does not hinge on the assumption of a closed (or semi-closed) economy but is, under certain circumstances, derived from the principle of comparative costs itself.

APPENDIX

PROPOSITION (1′): *Agricultural specialization is never optimum at points outside region A.*[13]

PROOF:

We have to prove that at least one optimality condition for agricultural specialization is violated at points to the left of E and above $\bar{k}_a(L)$ and to the right of E and above $\bar{k}_a(L)$. Take point Z of Figure 3. We first show that conditions (1) and (2) cannot hold simultaneously.

$$(1) \qquad \overset{\circ}{w}{}^{z} = p_a F_L^a(kL^z, L^z) \geqq q \frac{\partial F^b}{\partial L_b}(\overset{z}{k}_b)$$

$$(2) \qquad \overset{\circ}{r}{}^{z} = p_a F_K^a(kL^z, L^z) \geqq q \frac{\partial F^b}{\partial K_b}(\overset{z}{k}_b)$$

Suppose (2) holds. We then prove that (1) cannot hold. Consider point Z' on $\bar{k}_a(L)$ just below Z, i.e., $k^{z'} = k^z$; $L^{z'} < L^z$. Then,

$$(3) \qquad \overset{\circ}{r}{}^{z'} > \overset{\circ}{r}{}^{z}; \qquad \overset{\circ}{w}{}^{z'} > \overset{\circ}{w}{}^{z}$$

since reductions in scale increase the marginal products of both capital and labor in agriculture. Moreover, as Z' lies on $\bar{k}_a(L)$ we have (for $L_b = 0$)

$$\overset{\circ}{r}{}^{z'} = p_a F_K^a(kL^{z'}, L^{z'}) = q \frac{\partial F^b}{\partial K_b}(\overset{z'}{k}_b)$$

This relation together with (3) and (2) (which we assumed holds) implies

$$\frac{\partial F^b}{\partial K_b}(\overset{z'}{k}_b) > \frac{\partial F^b}{\partial K_b}(\overset{z}{k}_b) \text{ and hence } \overset{z}{k}_b > \overset{z'}{k}_b$$

Therefore

$$q \frac{\partial F^b}{\partial L_b}(\overset{z}{k}_b) > q \frac{\partial F^b}{\partial L_b}(\overset{z'}{k}_b) = \overset{\circ}{w}{}^{z'}$$

which, given (3), violates (1). Similarly, it can easily be proved that when (1) holds, (2) cannot hold.

PROPOSITION (2′): *In C_2 the optimum phase is heavy-industry specialization while in B_2 the optimum is light-industry specialization.*

PROOF:

We show that at points in C_2 the following optimality conditions for light-industry specialization cannot hold simultaneously:

[13] Under agricultural specialization the variables will be capped by °. Thus, $\overset{\circ}{k}_a = k$ and $\overset{\circ}{L}_a = L$. When the variables corresponding to a particular point in (k, L) are referred to, a superscript with the letter denoting the point will be added.

$$(4) \qquad \bar{r} = q \frac{\partial F^b}{\partial K_b} (\bar{k}_b) \geq \frac{\partial F^c}{\partial K_c} (k_c)$$

$$(5) \qquad \bar{w} = q \frac{\partial F^b}{\partial L_b} (\bar{k}_b) \geq \frac{\partial F^c}{\partial L_c} (k_c)$$

We start by assuming that (4) holds with equality in C_2, thus defining for each point a particular value of k_c which we call $\bar{k}_c$,

$$(4') \qquad \bar{r} = q \frac{\partial F_b}{\partial K_b} (\bar{k}_b) = \frac{\partial F^c}{\partial K_c} (\bar{k}_c)$$

We now show that given (4′), (5) cannot hold in C_2. Note that the C_2 region is bounded by the $\bar{k}_a(L)$ and $k_c(q)=k'$ schedules. Along $\bar{k}_a(L)$ (to the right of E) the optimum policy is heavy-industry specialization, which, given (4′), implies violation of (5). Along $k_c(q)=k'$, on the other hand, the optimum also involves heavy-industry specialization, although it is a border case where (5) holds with equality. Therefore, for (5) to hold with strict inequality at some point in C_2 given (4′), it must hold with equality at some other point in C_2, since all the functions in (5) are continuous. But, given (4′), (5) can only hold with equality in the non-specialization region, since in any other region, the nonnegativity restrictions would be violated. Therefore, given (4′), equation (5) cannot hold in C_2, i.e.,

$$(6) \qquad \bar{w} = q \frac{\partial F^b}{\partial L_b} (\bar{k}_b) < \frac{\partial F^c}{\partial L_c} (\bar{k}_c)$$

Suppose now that (4) holds with strict inequality. Then $k_c > \bar{k}_c$ and

$$\frac{\partial F^c}{\partial L_c} (\bar{k}_c) < \frac{\partial F^c}{\partial L_c} (k_c)$$

Together with (6), this violates (5). We conclude that heavy-industry specialization is optimum in C_2. An analogous analysis would prove that light-industry specialization is optimum in region B_2.

REFERENCES

A. Gerschenkron, *Continuity in History and Other Essays*, Cambridge, Mass. 1968.

R. W. Jones, "Factor Proportions and The Heckscher-Ohlin Theorem," *Rev. Econ. Stud.*, 1956, *24*, 1–10.

T. M. Rybczynski, "Factor Endowment and Relative Commodity Prices," *Economica*, Nov. 1955, *22*, 336–41.

M. Teubal, "Development Strategy for a Medium Sized Economy," *Econometrica*, Sept. 1971, *39*, 773–95.

The Informational Efficiency of Monetary Exchange

By JOSEPH M. OSTROY*

The gains from exchange may be extended if individuals are not required at each trade to balance the value of purchases and sales. But self-interest, the motive force of trade, does not move individuals to realize these gains. Suppose that X has an excess demand for ten apples and an excess supply of ten oranges. It may be efficient for X to receive the ten apples from Y, who has an excess supply, and later give his ten oranges to Z, who has an excess demand. In the actual execution, what is to prevent X from asking for eleven apples from Y, justifying his claim by saying he will supply an equal number of oranges, or give up only nine oranges to Z, saying that he took only nine apples? The purpose of this paper is to show that the essential property of money is to discourage the making of such inconsistent claims without also discouraging efficient patterns of trade.

This is "old stuff"; but it cannot be incorporated into the standard theory of value.[1] To illustrate, consider a paradox contained in Robert Clower's (1967) proposal to introduce money into the standard theory. His suggestion was to add to the existing budget constraint the injunction that current purchases be financed by sale of money only, not by current supplies of other commodities. This modification contradicts the belief that the introduction of money improves the allocation of resources. Because it is an additional constraint, it is at best not binding; and if binding, it will narrow the set of permissible exchanges compared to its barter counterpart. Clower's constraint makes no sense in the Walrasian model of exchange, but for a perfectly sensible reason. It does no good to append what is a trading constraint to a model which ignores trade.

In the standard theory, equilibrium is said to exist when (W) the sum of individual excess demands is zero for each commodity. What (W) defines is a Walrasian equilibrium of prices. Nevertheless, from the individuals' points of view, they are in disequilibrium as long as we do not have (A) all individual excess demands are zero for each commodity.

Should the economy reach the state defined by (W), will it then go to (A)? Interpretations of the standard theory say that the Walrasian auctioneer, after announcing equilibrium prices, expedites demands and supplies. If there is one theme which distinguishes the present treatment from the standard theory, it is that exchange is a do-it-yourself affair. Individuals will not exchange with "the market"; they will exchange with each other. This elementary logistical consideration is the basis upon which I shall construct an argument for monetary exchange.[2]

* Assistant professor of economics, University of California, Los Angeles. I have benefited from discussions with Ronald Britto, Jack Hirschleifer, Ben Klein, Axel Leijonhufvud, Louis Makowski, Richard Sweeney, Earl Thompson, and Joseph Wharton. Bryan Ellickson, Peter Howitt, and Ross Starr made many improvements in the formulation and presentation. My introduction to the topic came through the research of Robert Clower whose continued advice and encouragement were invaluable.

[1] By this phrase I mean the model of general equilibrium first proposed by Walras and recast into its definitive mathematical form by such contributors as Arrow and Debreu.

[2] Recently, formal notice has been taken of the fact that the valuable exchange services rendered by the

I. Summary

I shall assume that trade occurs between pairs of individuals so that the advantages of multilateral exchange must be obtained through a sequence of bilateral trades. During a unit interval of time, an individual meets with only one other so that during the interval an individual's trades are limited by his own and his trading partner's current endowments. When one pair meets, other pairs are also making contact, so that exchange occurs as a sequence of *simultaneous* bilateral trades.[3]

In the comparison of money and barter trading arrangements, no explicit account is taken of the physical or psychic costs of exchange. The single criterion is the number of periods it takes to accomplish the task of going from (W) to (A). Any decrease in the number is good and any increase is bad. Obviously a crude picture of the costs of exchange, it may be tolerated because the object is only an exposition of the role of money, not a general theory of trade.[4]

In the analysis of the problem of going from (W) to (A), I shall focus on the following three properties of trading sequences: (i) their technical feasibility, (ii) their informational feasibility, and (iii) their equilibrium properties, or what might be called their "behavioral feasibility." Property (i) defines the restrictions on a sequence of exchanges imposed by the fact that trade occurs between pairs. Property (ii) requires (i) and the restriction that a pair cannot base its trading decision on information available only to other pairs. For example, a pair cannot make its decisions depend on the full details of excess demands among all other pairs. Property (iii) requires (ii) and the restriction that each individual have no incentive to depart from the sequence. In a barter economy, if trades leading from (W) to (A) do not satisfy bilateral balance (BB)—where the value of purchases and sales are equated at each bilateral encounter—those trades will not form an equilibrium sequence.

We shall see that those trades which are technically and/or informationally feasible and which also minimize the number of periods in going from (W) to (A) will not satisfy BB. Imposition of BB does not preclude equilibrium in the sense of (A); it only means that it will take longer. But this time is not well spent because there is no technical or informational constraint underlying it. Additional time is required because individuals do not feel constrained to balance their budgets over a sequence of trades if they are not so compelled at each

auctioneer are costly to provide. Much of the motivation for these studies has been a desire to fit monetary exchange into the standard theory. We may learn from some of them that costly exchange can be introduced without giving up the assumption that exchange is coordinated by a central agency—an auctioneer who charges for his services. See Frank Hahn (1971), Mordecai Kurz, Jürg Niehans (1971), K. C. Sontheimer and N. Wallace. According to the present treatment, however, it is only when the exchange process is decentralized that the role of money can be understood. See Clower (1971). We may learn from Niehans (1969) that even when exchange is restricted to pairs it need not be completely decentralized. The selection of a least cost bilateral trading network can be made by a central planner who solves a complicated programming problem. A similar difficulty occurs in Starr (1970) where individuals choose optimal *sets* of bilateral transactions but require a central agency to hook them together. Roy Radner and Karl Brunner and Allan Meltzer have approached monetary exchange as a reflection of imperfect information. Radner has suggested that money might arise from the unpredictability of future spot prices and Brunner and Meltzer have indicated that money arises because of the need for a commonly recognizable asset. I shall discuss these points in Section IV. The present treatment is related to the work of Hicks, Starr (1972), and E. C. H. Veendorp.

[3] It is this simultaneity—while one pair is exchanging other pairs are not standing still—which contributes to the informational demands on trade.

[4] It would have been possible to formulate a model in which the costs of exchange varied with the amounts of commodities and the number of individuals with whom one traded per unit time. Ignoring the resulting complexities, the outcome must concede that not everyone exchanges everything at once if it is to gain insight into monetary exchange.

trade. Any device which would encourage such constraint could be substituted for the added time. Money is such a device.

I shall assume throughout that individuals know equilibrium exchange rates. Because we are accustomed to thinking of equilibrium in the sense of (W), this assumption may appear to be disquietingly strong. In general, it is; but *not* for the purposes of understanding monetary exchange. The line of reasoning adopted here permits me to assert that if we cannot find a role for money when equilibrium prices are known, we shall not find one when they are unknown. There can hardly be a speculative demand for the medium of exchange without a transactions demand and this transactions demand does not depend on price uncertainty.

II. A Model of a Trading Economy

There are three components of the model: (i) the set of possible tastes and endowments and their corresponding competitive equilibrium allocations; (ii) the logistical description of the trading arrangement; and (iii) the pattern of information describing what individuals know and do not know at each trading opportunity.

A Family of Exchange Economies

Individual i is initially endowed with a nonnegative quantity w_{ic} of commodity $c=1, \ldots, m$. The complete list of his initial endowments is the m-vector $w_i=(w_{i1}, \ldots, w_{im})$. If i, whose tastes are represented by $u_i(\cdot)$, were asked to exchange his initial endowment for any other (m-vector) x_i satisfying the constraint $p\cdot x_i=p\cdot w_i$, where $p=(p_1, \ldots, p_m)$ is the vector of prices, let his answer be the bundle $a_i=(a_{i1}, \ldots, a_{im})$, $a_{ic}\geq 0$, $c=1, \ldots, m$ and

$$(1) \qquad u_i(a_i) = \max u_i(x_i), \quad p\cdot a_i = p\cdot x_i = p\cdot w_i{}^{5}$$

[5] $p\cdot w_i\equiv\Sigma_c p_c w_{ic}$ is the dot product of p and w_i.

At prices p, let i be described by the pair of vectors (a_i, w_i) and let the collection of individuals $i=1, \ldots, n$ be described by the pair of matrices (A, W), where a_i and w_i are the ith rows of the nxm matrices A and W, respectively. The pair (A, W) forms a competitive equilibrium (CE) if the aggregate demand for each commodity is equal to the aggregate supply,

$$(2) \qquad \sum_i a_{ic} = \sum_i w_{ic}, \qquad c = 1, \ldots, m$$

The matrix A *is the CE allocation* for the CE price vector p and matrix of initial endowments W. The problem of going from equilibrium in the sense of (W) to equilibrium in the sense of (A) is now reduced to the problem of going from the matrix W to the matrix A.

I shall deal only with collections of individuals described by nonnegative matrices (A, W) satisfying

$$(3) \qquad n = m$$

and

$$(4) \qquad \sum_i a_{ic} = \sum_c a_{ic} = \sum_i w_{ic} = \sum_c w_{ic} = 1$$

The set of all such pairs of nxn matrices whose row and column sums are unity, call $\mathcal{U}$. If (4) is to satisfy (1) for every member of $\mathcal{U}$, any $p=(r, r, \ldots, r)$, $r>0$, must be a CE price vector. This will mean that if i knows he is in an economy belonging to $\mathcal{U}$, he knows equilibrium exchange rates.

In the space of all possible economies, the set $\mathcal{U}$ occupies only a small corner. I choose to deal with it because $\mathcal{U}$ exhibits the salient features of the general case. The assumption in (3) that the number of commodities is equal to the number of individuals is, strictly speaking, not essential to our results and can be shown to follow a fortiori if $m>n$. However, the trading ar-

rangement I shall postulate allows each individual's trading opportunities to increase with the size of the population so that as n/m increases the logistical problems of exchange disappear.

The assumption that individual endowments are of the same size ($\sum_c w_{ic}=1$, $i=1, \ldots, n$) is significant only insofar as it makes clear that there is no individual whose initial endowment is, for *all* $(A, W) \in \mathcal{U}$, large enough for him to act as a central distributor supplying everyone else's excess demands.[6] The set $\mathcal{U}$ exhibits a similar feature with respect to the insufficiency of endowments to permit a medium of exchange.

About the origin and use of money, Adam Smith said:

> In order to avoid the inconvenience of such situations, every prudent man in every period of society, after the first establishment of the division of labour, must naturally have endeavoured to manage his affairs in such a manner, as to have at all times by him besides the peculiar produce of his own industry, a certain quantity of some one commodity or other, such as he imagined few people would be likely to refuse in exchange for the produce of their industry. [p. 22]

The purpose of this paper is to isolate the function of a medium of exchange and I shall proceed by analyzing the difficulties when one does not have "a certain quantity of some one commodity." There is no commodity whose initial value is, for *all* $(A, W) \in \mathcal{U}$, a significant fraction of the value of each individual's planned purchases.[7] This will allow me to bring out more clearly that the essential feature of monetary exchange has its origin in the trading arrangement and not in the nature of the money commodity.[8]

The first result, which serves as an introduction to the problem, is

PROPOSITION 1: *For almost all $(A, W) \in \mathcal{U}$, if the collection of n individuals is divided into any two groups consisting of k and $n-k$ individuals, $1 \leq k \leq n-1$, who cannot trade with each other, the CE allocation cannot be obtained.*[9]

If $n>2$, there is little hope of finding a double coincidence of wants. In fact, everyone will have to depend on everyone else if the *CE* allocation is to be realized.

How Traders Meet

I shall assume that the sequence of pairwise meetings is parametric rather than a subject for choice. From his point of view, each individual seems to collide every so often with someone else.

Let $\pi = \{\pi^t\}$, $t=1, \ldots, \tau$ be a sequence of permutations of the integers $i=1, \ldots, n$ such that for all i and t,

$$\pi^t(i) = j \tag{5}$$

if and only if $\pi^t(j) = i$

The permutation π^t determines who meets whom at t; i.e., $\pi^t(i) = j$ means i and j are trading partners at t. The final period, after which all trading ceases, is $t=\tau$.

Let $\{W^t\}$, $t=1, \ldots, \tau+1$, τ be a sequence of matrices with nonnegative elements where $W^t = (w^t_{ic})$ is the matrix of endowments at the beginning of t. Suppose $\pi = \{\pi^t\}$ describes the sequence of meetings; then $\{W^t\}$ *is technically feasible for* π, if for all i and t and $\pi^t(i) = j$,

$$\overset{t+1}{w_i} + \overset{t+1}{w_j} = \overset{t}{w_i} + \overset{t}{w_j} \tag{6}$$

[6] For example, individual j could be a central distributor for (A, W) if $w_{jc} > \delta$, $c=1, \ldots, m$, and if max $|a_{ic} - w_{ic}|$ over all c and $i \neq j$ were less than δ/n.

[7] For example, commodity d could be used as a medium of exchange for (A, W) if $w_{id} > \delta$, $i=1, \ldots, n$, and if max $|a_{ic} - w_{ic}|$ over all i and $c \neq d$ were less than δ/m.

[8] These qualifications as well as their consequences are more fully discussed in Ostroy and Starr.

[9] PROOF: Suppose the contrary; then it would require that for some subset, T, consisting of fewer than n individuals, $\Sigma_{i \in T}(a_i - w_i) = 0$. This defines a less than full dimensional class of economies in $\mathcal{U}$.

This says that an individual can add to his current endowment only by subtracting from the current endowment of his current trading partner and that commodity totals are not changed in the process of trade, just redistributed.

If the sequence $\{W^t\}$ *also* satisfies

$$(7) \qquad p\cdot(w_i^{t+1} - w_i^{t}) = 0$$

for all i and t, where p is the *CE* price vector, then *bilateral balance* (*BB*) obtains. At every bilateral encounter, the value of what is given up is equal to the value of what is received. Clearly, *BB restricts the set of trades beyond the demands of technical feasibility.*

Let us agree to say that the *CE allocation is technically feasible for* π if, for *all* $(A, W)\in\mathfrak{U}$, there exists a sequence $\{W^t\}$, technically feasible for π, with $W^1=W$ and $W^{\tau+1}=A$. This means that the sequence π is not biased; it permits all possible *CE* configurations to be realized. Denote by Π_τ the set of all such π of length τ.

What is the minimum value of τ for which the *CE* allocation is technically feasible? Proposition 1 tells us that everyone must be "connected" to everyone else but this connection need not be direct since (6) permits indirect or middleman trade.

PROPOSITION 2: *If* $n=2^k$, $k=1, 2, \ldots,$ *the minimum number of periods for which the CE allocation is technically feasible is*[10]

$$\tau = k(\tau = log_2\, n)$$

To demonstrate, note that it is true for $k=1$, and assume it is true for $k=q$. This means that any group consisting of 2^q individuals can be connected in q periods so that two groups each consisting of 2^q individuals can be connected in period $q+1$ if every member of the one group is assigned to a member of the other. Since $2\cdot 2^q=2^{q+1}$, Proposition 2 is proved.

Call such a π which satisfies Proposition 2 an *indirect trading sequence* and a trading economy which makes use of it an *indirect exchange model.* It should be pointed out that

PROPOSITION 3: *If trades must satisfy BB, the CE allocation is not technically feasible for the indirect exchange model.*[11]

Suppose every individual is permitted to meet every other directly. Obviously, the *CE* allocation will be technically feasible. This will require $n(n-1)/2$ bilateral meetings and assuming n is even, will take, at a minimum, $\tau=n-1$ periods. Call such a sequence which allows everyone to meet everyone else in a minimum number of periods a *direct exchange sequence* and a trading economy which makes use of it a *direct exchange model.* An interesting feature of this model is

PROPOSITION 4: *If trades must satisfy BB, the CE allocation is technically feasible for the direct exchange model.*[12]

Propositions 3 and 4 say that it takes longer to reach the *CE* allocation if *BB* is imposed. Alternatively, the temporal advantages of the indirect exchange model are incompatible with *BB*. This will not, by itself, offer a basis for monetary exchange. Other considerations intrude.

How Trades Are Made

I shall make a distinction between (a)

[10] The following was obtained in correspondence with Lloyd Shapley: To complete the solution to minimizing the number of time periods for $n\neq 2^k$, let m be the largest integral power of two such that $m=2^k<n$. Then, $\tau=(log_2 m)+1$ if n is even and $n\neq 2^k$ and $\tau=(log_2 m)+2$ if n is odd.

[11] Proposition 3 holds for all $n\geq 4$, the smallest number for which the advantages of indirect trade appear. See the *Remark* following Proposition 5 for a demonstration in the case $n=4$.

[12] Proofs of Proposition 4 are in Ostroy and Starr, and G. Bradley.

the decision a pair of traders make based on the information they reveal to each other and (b) the decision as to what information to reveal. It will be assumed that (a) is taken out of the hands of the pair and given to a fictional third party, a broker, who makes the trading decision solely on the basis of what each member of the pair tells him. We may suppose that the broker uses his unlimited ingenuity and computational capacity to help the pair reach their *CE* allocations. His only constraint is that he knows no more than what the pair tells him. Until Section IV, the decision (b) will be ignored by assuming that there is no distinction between what individuals know and what they reveal.[13]

At the beginning of period t, those features of the economy which cannot be changed constitute the state of the economy. They are the initial configuration of tastes and endowments, the order in which pairs will meet, and the trades which have been made up to t. Denote a typical state by $\hat{S}^t=(\hat{A}, \hat{W}^1, \ldots, \hat{W}^t; \hat{\pi})$. The set of all possible states at t is

$$(8) \quad \mathcal{S}^t = \{\hat{S}^t: (A, \hat{W}^1) \in \mathcal{U} \text{ and } \{\hat{W}^k\}, \; k = 1, \ldots, t \text{ is technically feasible for some } \hat{\pi} \in \Pi_\tau\}$$

Suppose the actual state is $S^t=(A, W^1, \ldots, W^t; \pi)$. Let $I_i^t(S^t)$ be the information i has at t about the actual state. It will be assumed throughout that

$$(9) \quad I_i^t(S^t) = \{\hat{S}^t \in \mathcal{S}^t: \hat{a}_i = a_i, \hat{w}_i^k = w_i^k, \; \hat{\pi}^k(i) = \pi^k(i), k = 1, \ldots, t\}$$

If you were to ask i what he knows about the state of the economy, according to (9) he would say: "I have no idea. All I know is that it belongs to the set of possible states, and I can therefore tell you what *CE* prices are, and what I would want if I had to balance my budget at those prices, and that I have made certain trades as indicated by $\{w_i^k\}$, $k=1, \ldots, t-1$, leading to my current position w_i^t."[14]

Each member of a pair of trading partners tells what he knows to the broker who then decides what trades they should make. The different trading pairs have different brokers who cannot communicate so the situation is much the same as if the pair themselves decided what to trade.

As a formalization of this story, let $\rho_i^t(I_1^t(S^t), \ldots, I_n^t(S^t))=w_i^{t+1}$, $i=1, \ldots, n$, be a trading rule, changing w_i^t into w_i^{t+1}, which depends on the actual state of the economy and what individuals know about it. I shall say that $\rho=\{\rho_i^t\}$ *is an informationally feasible trading rule* if for all i and t, and $\pi^t(i)=j$,

$$(10) \quad \rho_i^t + \rho_j^t = w_i^t + w_j^t$$

and, for all $\hat{S}^t \in I_i^t(S^t) \cap I_j^t(S^t)$,

$$(11) \quad \rho_i^t(I_1^t(\hat{S}^t), \ldots, I_n^t(\hat{S}^t)) = \rho_i^t(I_1^t(S^t), \ldots, I_n^t(S^t))$$

Condition (10) says the rule must be technically feasible. If the pair (i, j) were to share their information they could determine that the actual state was in the set $I_i^t \cap I_j^t$ and nothing more. Condition (11) says that the trading decision must respect this ignorance, which is to say that each pair's trading decision cannot be made contingent on the tastes and trading histories of other pairs.[15]

[13] Getting individuals to reveal what they know, for example, their tastes, has been recognized as the principal difficulty in allocating collective goods (see Paul Samuelson). We shall see that there are similar strategic issues in a barter economy.

[14] This is the game-theoretic method for describing imperfect information. J. C. Harsanyi has shown how this may be applied to the case of players in a game who do not know each others' payoffs. The treatment above was developed independently in Ostroy (1970).

[15] The less one knows the smaller is his set of possible strategies. See Radner for a restriction similar to (11).

Once a trading rule is selected, the course of the economy is uniquely determined by its initial state. Given $S^1=(A, W^1; \pi)$, (9) determines $\{I_i^1(S^1)\}$, the input into the trading rule $\{\rho_i^1\}$, which determines W^2 and therefore $S^2=(A, W^1, W^2; \pi), \ldots$, etc. To summarize this recursive relation, let us say that if the initial state is $(A, W; \pi)$ and the trading rule is ρ, the end result is $g_i[\rho|(A, W; \pi)] = w_i^{\tau+1}$, $i=1, \ldots, n$.

Now, *the CE allocation is informationally feasible* if there exists an informationally feasible trading rule $\rho=\{\rho_i^t\}$ such that for *all* $(A, W)\in\mathfrak{U}$ and $\pi\in\Pi_\tau$, $g_i[\rho|(A, W; \pi)] = a_i$, $i=1, \ldots, n$. To illustrate this definition, consider the first period trading decision for any pair (i, j). If the CE allocation is informationally feasible, their broker has a sure-fire method for putting them on a path leading to their CE allocation no matter what the values of a_k and w_k, $k\neq i, j$.

III. Informational Aspects of Trade

In this section, the consequences of informational feasibility are explored. First, we have

PROPOSITION 5: *The CE allocation is not informationally feasible in the indirect exchange model.*

To demonstrate, take $n=4$. Let the trading partners be assigned as follows: individual 1's trading partners in the first and second periods are $\pi^1(1)=2$ and $\pi^2(1)=3$ while 4 is the partner of 3 in the first period and of 2 in the second $-\pi^1(3) = \pi^2(2)=4$. It is readily verified that this sequence allows each individual to trade directly or indirectly with everyone else. From Proposition 2, this is the minimum number of periods since $\tau=log_2 4=2$. Assume that initial endowments are given by the identity matrix; i.e., $w_{ic}=1$, if $i=c$, and $w_{ic}=0$ if $i\neq c$, where $i, c=1, 2, 3, 4$. This will simplify the demonstration, but it is not essential.

To go from W to A it is necessary that the exchange between individuals 1 and 2 be such that 1 begin the second period with $w_1^2=(w_{11}^2, w_{12}^2, w_{13}^2, w_{14}^2)$ where

$$(12) \quad w_{1c}^2 = \begin{cases} 1-(a_{2c}+a_{4c}), & \text{if } c=1 \\ (a_{1c}+a_{3c}), & \text{if } c=2 \\ 0, & \text{otherwise} \end{cases}$$

Individuals 1 and 2 do not know the tastes of 3 and 4, given by the vectors a_3 and a_4. According to (11), this means that for all possible values of a_3 and a_4, 1 and 2 must make the same trade. Now, whatever trade they make, they will have made the right decision for at most one pair (a_3, a_4) and will have made the wrong decision in all other cases. Therefore, the CE allocation is not informationally feasible.

Remark. The rule (12) is compatible with BB only if the configuration of tastes satisfies $a_{21}+a_{41}=a_{12}+a_{32}$; i.e., almost never.

I have assumed that individuals know only equilibrium prices and their own tastes and endowments. They do not know each other's excess demands and trading decisions are bound by this ignorance. But, to take advantage of indirect exchange, individuals must act as middlemen passing excess supplies in just the right sequence of intermediary trading to the final demander. What is the right sequence depends on the entire configuration of excess demands as well as on the order in which all pairs will meet. Proposition 5 brings out an obvious and basic point: *the informational requirements for indirect trade go beyond a knowledge of prices.*

Of course, the restrictions imposed on $\{I_i^t\}$ are rather harsh. I have taken the information available to an individual in the standard theory where he does not have to do his own trading and inquired as

to its sufficiency where he does. From here, it would be possible to go on to find the minimum information compatible with the result that the *CE* allocation is informationally feasible in the indirect exchange model. It has already been determined from Propositions 3 and 4 that the required trades will not satisfy *BB*. Therefore, the argument could be made that only with money would the individuals reveal what they knew. I shall not follow this course because the informational demands for the indirect exchange model appear to be complicated and are certainly exorbitant. The same argument can be more easily elaborated with the direct exchange model.[16]

PROPOSITION 6: *If trades must satisfy BB, the CE allocation is not informationally feasible in the direct exchange model.*[17]

The reasons for Proposition 6 are similar to those underlying Proposition 5. In the class of economies $\mathfrak{U}$, no commodity is in sufficient supply to serve merely as a balancing item; and it will not do to pay one's debts in just any commodity or commodities. If the *CE* allocation is to be achieved under *BB*, commodities used as payment by i in his trade with j must also be those which j can pass on to $k, \ldots,$ etc., so that they pass in just the right sequence and end up in just the right hands. But this involves the individuals in indirect trade whose informational demands they cannot meet.

Proposition 6 points to a feature of the medium of exchange distinguishing it from a standard I.O.U. *When money is used, the parties to the transaction are admitting their inability to predict with whom and how the account will be settled.*

Proposition 6 is important because of its relation to

PROPOSITION 7: *When BB is not imposed, the CE allocation is informationally feasible in the direct exchange model.*

The trading rule which demonstrates Proposition 6 is: for all i and t, $\pi^t(i)=j$, and $\rho_i^t=(\rho_{i1}^t, \ldots, \rho_{im}^t)$, let $\rho_{ic}^t, c=1, \ldots, m$ be such that

(13)

$$\rho_{ic}^t=\begin{cases} w_{ic}^t+\min\left[\,|a_{ic}-w_{ic}^t|,\ |a_{jc}-w_{jc}^t|\,\right] \\ \qquad \text{if } (a_{ic}-w_{ic}^t)\geq 0 \text{ and } (a_{jc}-w_{jc}^t)<0 \\ w_{ic}^t-\min\left[\,|a_{ic}-w_{ic}^t|,\ |a_{jc}-w_{jc}^t|\,\right] \\ \qquad \text{if } (a_{ic}-w_{ic}^t)<0 \text{ and } (a_{jc}-w_{jc}^t)\geq 0 \\ w_{ic}^t, \text{ otherwise} \end{cases}$$

The rule described by (13) is an example of what Ross Starr (1972) has called excess demand diminishing (*EDD*) trades. They follow the prescription "never engage in any trade which changes the sign of your excess demand." If you start out as a buyer of a commodity, do not accept more than you planned to purchase; and, if you start out as a seller of a commodity, do not give more than you planned to sell. The prescription is designed to prevent indirect or middleman trade.

It is clear that (13) satisfies the technical feasibility condition (6) and is informationally feasible since it requires only a knowledge of the trading pair's current excess demands, $(a_i-w_i^t)$ and $(a_j-w_j^t)$. It is a straightforward matter to show that (13) must always result in the *CE* alloca-

[16] The indirect exchange sequence, by explicitly denying the informational sufficiency of prices, provides the kind of environment hospitable to the activities of specialists in exchange. See Ostroy (1970). All of this is of a piece with money. Nevertheless, when our interest is just monetary exchange, we may use the direct exchange model to isolate the essential aspects of the problem in the context of a neater solution.

[17] Proofs of Proposition 6 follow along the lines of the demonstration of Proposition 5. Clearly, for $n=1$ and 2, it is false. It is also false for $n=3$. The reason is that once any two information sets $\{I_i^t\}$, $i=1, 2, 3$, are known the other may be inferred so there is effectively perfect information. Unless the additional assumption is made that individuals do not know their own trading histories, Proposition 6 is false for $n=4$. See Ostroy and Starr for a proof when $n\geq 5$.

tion if everyone is able to meet everyone else.

The merits of *EDD* trades are substantial. In terms of information, they are extremely economical and they lead, in the direct exchange (but not the indirect exchange!) model, to the *CE* allocation. They have demerits as well.

1. *EDD trades do not satisfy BB*. This is to be expected in light of Proposition 6. For emphasis, I shall add that they almost never satisfy *BB* (if $n>2$). To illustrate, consider the case of $W=I$, the identity matrix. Then, for all $(A, I)\in\mathfrak{U}$, *EDD* trades satisfy *BB* if, and only if, A is symmetric.[18]

2. *EDD trades do not form a utility increasing sequence.* Whenever $p\cdot(w_i^{t+1}-w_i^t)<0$, so that sales exceed purchases, we must admit the possibility that $u_i(w_i^{t+1})<u_i^t(w_i^t)$. Similarly, whenever $p\cdot(w_i^{t+1}-w_i^t)>0$, so that purchases exceed sales, we may have $u_i(w_i^{t+1})>u_i(w_i^t)$; and, if some of one's purchases are made before any of one's sales, we may have $u_i(w_i^t)>u_i(a_i)$! This agrees with everday experience. If you did not have to pay for your purchases your utility would be above what it otherwise is.

3. *EDD trades are unpredictable.* If an individual has a positive (negative) excess demand at the start of t, he cannot tell how much of it will be fulfilled (taken) during the period. To know this, he would have to know the entire configuration of initial excess demands as well as who met whom before t. This means that during the course of trade, an individual cannot determine whether he is on a path leading to his *CE* allocation or some other point. Only at $t=\tau+1$ does he know where he ends up. Suppose i does not end up at a_i because, to jump ahead, someone other than i misrepresented himself. Now, i knowns this was not his fault, but he cannot determine from his trading positions, $w_i^1, w_i^2, \ldots, w_i^{\tau+1}$, who was responsible. He cannot, for example, surmise that if $\pi^t(i)=j$, and $p\cdot(w_i^{t+1}-w_i^t)<0$, that j overstated his demands. It may have been that at period s, when $\pi^s(i)=k$, and $p\cdot(w_i^s-w_i^s)>0$, that individual k understated his supplies.

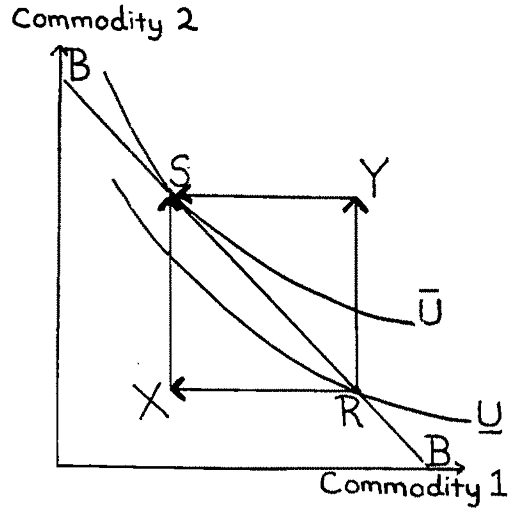


FIGURE 1

The above three features of *EDD* trades are illustrated in Figure 1. The straight line, *BB*, is the budget line. It goes through the initial endowment, point *R*, and the *CE* allocation, point *S*. The *EDD* trades begin and end on the budget line but in the interim they depart from it. The path $R\rightarrow X\rightarrow S$ indicates a trading sequence in which sale preceded purchase while the path $R\rightarrow Y\rightarrow S$ indicates the reverse. There is no guarantee that the intermediate posi-

[18] Applying (13) when $W=I$, we have that for any i and $\pi^t(i)=j$,

$$\rho_{ic}^t=\begin{cases} w_{ic}^t+a_{ic}, & \text{if } c=j \\ w_{ic}^t-a_{jc}, & \text{if } c=i \\ 0, & \text{otherwise}\end{cases}$$

Therefore, when $p=(r, r, \ldots, r), r>0, p\cdot(w_i^{t+1}-w_i^t)=0$ if, and only if, $a_{ij}=a_{ji}$.

tions, indicated here by the points X and Y, will lie within the region whose lower bound is $\underline{U}$, the indifference curve passing through the initial endowment, and whose upper bound is $\overline{U}$, the indifference curve passing through the CE allocation.

IV. Equilibrium

Recall that in defining the trading rule, $\rho=\{\rho_i^t\}$, it was assumed that individuals had no choice. Each revealed what he knew to the best of his knowledge. However, the trading rule does not require this accuracy. All that is required and all I shall assume is that no one says he has more of a commodity than he knows he has, although he may say he has less.

Let $h_i^t(I_i^t)=\hat{I}_i^t$ be the information i conveys to the broker. It is his decision whether or not to misrepresent what he knows. A strategy for i is a system $h_i=\{h_i^t(I_i^t)\}$, detailing what he will say he knows given what he actually knows at each trading opportunity. Therefore, $\{h_i^t\}$ rather than $\{I_i^t\}$ will be the informational inputs determining the course of trade.

Suppose a trading rule $\rho=\{\rho_i^t\}$ and strategies h_i^t, $i=1,\ldots,n$ have been selected. The outcome is uniquely determined by the initial state. If $S^1=(A, W^1; \pi)$, this determines $\{I_i^1(S^1)\}$ and therefore $\{h_i^1(I_i^1)\}$, and then $\{\rho_i^1(h_1^1,\ldots,h_n^1)\}$ which determines W^2 and therefore $S^2=(A, W^1, W^2; \pi),\ldots,$ etc. We may summarize this recursive relation by saying that if the initial state is $(A, W; \pi)$ and the trading rule is given by $\rho=\{\rho_i^\tau\}$ and individual strategies are given by $h_1,\ldots,h_n$, the end result is

$$g_i[\rho, h_1,\ldots,h_n \mid (A, W; \pi)] = w_i^{\tau+1}, \qquad i=1,\ldots,n$$

Now, we may say that *the CE allocation is an equilibrium* if for *all* $(A, W)\in\mathfrak{U}$ and $\pi\in\Pi_\tau$ there exists a trading rule $\rho^*=\{\rho_i^{*t}\}$ and strategies $h_1^*,\ldots,h_n^*$ such that for all $i=1,\ldots,n$,

$$(14)\qquad g_i[\rho^*, h_1^*,\ldots,h_n^* \mid (A, W; \pi)] = a_i$$

and

$$(15)\qquad u_i(a_i) = \max_{h_i} u_i(g_i[\rho^*, h_1^*,\ldots, h_i,\ldots,h_n^* \mid (A, W; \pi)])$$

The CE allocation is an equilibrium if the CE allocation is informationally feasible (14) and if it is in no individual's interest to depart from his strategy given the trading rule and strategies of the others (15).[19]

Underlying the question of the equilibrium of the CE allocation is the necessary condition of budget balance (BUB),

$$(16)\qquad p\cdot(w_i^{\tau+1} - w_i^1) = 0, \qquad i=1,\ldots,n$$

Over the course of trade, if not at each trade, (16) says the purchases and sales balance.

Supposing individuals to have no choice so that they are compelled to satisfy BUB, it follows immediately from the definition of the CE allocation (see (1)) that

PROPOSITION 8: *If the CE allocation is informationally feasible and if BUB is imposed the CE allocation is an equilibrium.*

If BUB were not imposed from outside, would individuals voluntarily choose strategies leading to it; i.e., would they choose to reveal what they know? If not, the CE allocation is not an equilibrium.

The only means the individuals have of imposing BUB amongst themselves is through BE. Suppose that all $i\neq j$ give instructions to their brokers that they will refuse to partake in any trade for which $p\cdot(w_i^{t+1}-w_i^t)<0$; then j cannot do other than satisfy BUB over the course of trade. If some i does not adhere to this, j is not

[19] Condition (15) is the definition of a noncooperative equilibrium proposed by John Nash.

constrained by *BUB*. It has already been determined (Proposition 6) that *BB* precludes the informational feasibility of the *CE* allocation in the direct exchange model. But to give up *BB* is to give up the incentive to reveal what one knows. The summary conclusion is

PROPOSITION 9: *The CE allocation is not an equilibrium for the direct exchange model.*[20]

Of course, if the individuals are willing and able to allow more time for trade, the opposite conclusion may be drawn. Consider the following "turn-taking" routine: Individual 1 goes to each of the others in turn asking for commodities to fulfill his (positive) excess demands and paying for them with his excess supplies so that *BB* is maintained. After $(n-1)$ periods he will reach his *CE* allocation. Next, 2 takes his with $3, \dots, n$ in the same manner, ..., etc., so that after $n(n-1)/2$ periods the *CE* allocation is achieved.[21]

V. Money as a Record-Keeping Device

How to enforce *BUB* without imposing *BB*? Rather than ask how this enforcement is actually effected, I shall focus on the conditions which must precede enforcement. Again, it is a matter of information.

Consider the direct exchange model and assume that if an individual who has "over-balanced" his budget—$p\cdot(w_i^{\tau+1}-w_i^1)>0$—is found out, he will be made to give up more than will put him back in balance; but if he is not found out, he may keep what he has. This represents a shortcut to the conclusions derived from a more extensive version (see Ostroy (1971)) which supposed that individuals were in an economy consisting of a large number of repetitions of the above direct exchange model.

At the completion of trade, we may ask i whether j balanced his budget. Recall from the definition of $I_i^{\tau+1}$, what he knows at the completion of trade, that i cannot say. Individual i knows only his own trading history and this will not suffice to infer what j has done. However, he does know one fact about j's trading history. At t, when $\pi^t(i)=j$, $p\cdot(w_j^{t+1}-w_j^t)=-p\cdot(w_i^{t+1}+w_i^t)$. If the information possessed by all $i\neq j$ were added together, we could compute $\sum_{t=1}^{t=\tau} p\cdot(w_j^{t+1}-w_j^t)=p\cdot(w_j^{\tau+1}-w_j^1)$. As it stands, however, this information is scattered among the individuals with no one other than j, himself, able to determine whether he has balanced his budget.

As a monetary version of the model of a trading economy, introduce a central receiving station called a monetary authority. Its function is to collect and collate the bits of information individuals have about each others' trading histories. Each will require his trading partner to write a signed statement, a check, indicating the amount by which the partner's purchases exceed his sales. This record is forwarded to the monetary authority who revises individual accounts on the basis of this new information. *Sellers, by requiring payment in money, are guaranteeing a steady flow of information such that the monetary authority, and it alone, is able to monitor trading behavior.* Of course, there is every incentive to require and deposit this information with the monetary authority; otherwise, one would not receive credit for sales and so have to cut back on purchases. Therefore,

[20] This result does not hold when $n<5$ (see fn. 17).

[21] This is an upper bound and not necessarily a minimum estimate of the number of periods required to make the *CE* allocation informationally feasible with *BB*. This result demonstrates that to achieve the same end additional time can be substituted for lack of information. The tradeoff may be pushed further. Allan Feldman has shown that when individuals know nothing but their own tastes and endowments but have an unlimited number of bilateral trading opportunities, if they accept only utility-increasing trades they will, under certain assumptions, eventually reach a Pareto optimal allocation.

PROPOSITION 10: *In the monetary version of the direct exchange model, the CE allocation is an equilibrium.*[22,23]

There is a small slip. We have seen that for any $(A, W) \in \mathfrak{U}$, any vector $p=(r, r, \ldots, r), r>0$, is a *CE* price vector. If the monetary authority is to be able to make trades between different individuals commensurable, they must all agree to the same value of r; i.e., we require a common unit of account. While this convention is essential to the operation of the record-keeping system, it is not identical to it. Money is not simply a unit of account.

VI. Conclusion: Integrating Monetary and Value Theory

In the Walras-Hicks-Patinkin tradition, the goal of monetary theory has been to present a picture of a money economy which would be a logical extension of the standard theory of value. Walras brought the equation of the offer and demand for money into line with the rest of his system by making a distinction between the stock of money, assumed to be without any utility of its own, and the "services of availability" of the stock which does contribute to one's well-being. Just as no inquiry is made into the sources of satisfaction from other goods, the services of availability are similarly unquestioned.

Recently, several theorists have suggested that it might be useful to dig a little more deeply.[24] While the traditional approach could be readily applied to determine why individuals might hold more or less money given that they valued it in certain ways in the first place, perhaps, just perhaps, it might clarify some contemporary monetary muddles if we asked why money is held at all.

This poses a dilemma. How to make money appear without making the standard theory disappear? Normal research strategy says that for a theory to be complete and consistent it must be derivable from the standard theory. But the standard theory has been cultivated to its present high level as a model of exchange in which money does not appear. Unlike the Walras-Hicks-Patinkin approach which left the standard theory intact and relied on conceptual appendages to introduce money, the recent approach forces us to look for modifications within the body of the theory. The following are some suppositions as to where to look on which I have not relied. I shall argue that they are superseded by the conclusions obtained from the above model of a trading economy.

1. *Money enlarges the set of feasible transactions.*[25] In the standard theory, any redistribution of commodities which preserves their totals is feasible. Into this model we can introduce the problems of exchange as a kind of transport cost of getting from one bundle of goods to another. We may then reason that monetary exchange represents a least cost network, so that without the money commodity the set of feasible transactions must shrink. While this may be adequate metaphorically, it misses the point. It is fairly well

[22] PROOF: Let h_i^* be such that $h_i^{*t}(I_i^t)=I_i^t$, $i=1, \ldots, n$ and let ρ^* be the *EDD* trades given by (13). These verify the first condition, (14), that the *CE* allocation is technically and informationally feasible.

The *assumed* informational and enforcement powers of the monetary authority make it clear to each individual that he cannot spend more than he earns so that besides the technical and informational restrictions, he must obey $p \cdot (w_i^{\tau+1}-w_i^1) \leq 0$. Therefore, if $h_j^t(I_j^t) \neq I_j^t$ for some j and t, then $u_j(w_j^{\tau+1}) \leq u_j(a_j)$. Since it is in each individual's interest to reveal accurately his excess demands, the remaining equilibrium condition, (15), is satisfied.

[23] I have avoided use of the term "transactions costs" because its meaning varies from writer to writer. However, the present treatment seems consistent with the the usage of my colleague, Earl Thompson, who defines transactions costs as those losses arising from differences in information. With his definition, we may say: Money reduces transactions costs.

[24] Compare fn. 2. That we recognize there is a problem at all is due largely to the important papers of Don Patinkin, Jacob Marschak, Hahn (1965), and Clower (1967).

[25] See Hahn (1971), Kurz, Niehans (1971) and Sontheimer.

established that the term "feasible" denotes what *could* happen, ignoring individual behavior, not what *would* happen. Monetary exchange does not enlarge the set of feasible transactions; it merely enables trades, which must be feasible in the first place to be realized.

2. *Money is held because we do not know what prices will be.*[26] Price uncertainty is neither necessary nor sufficient to explain the presence of a medium of exchange. The above model of a trading economy assumed that exchange rates were known; yet it required a record-keeping device. Suppose, however, that exchange rates were unknown but that individuals voluntarily agreed to keep accurate records of their transactions (in terms of a common unit of account) in order to balance their budgets. There would be no need for a medium of exchange.[27]

3. *The advantages of money have their origin in the properties of the money commodity.*[28] In the above model of a trading economy, all commodities are perfectly portable, durable, divisible, and recognizable, yet there is a need for money. The origin of this need is the decentralized trading arrangement. I chose to introduce a monetary authority and bookkeeping entries as a kind of ideal monetary arrangement because the record-keeping function of money is conceptually distinct from the properties of the commodities traded. Of course, to understand a particular monetary arrangement, it becomes a matter of recognizing a minimum cost method of imposing budget balance and in a society unfamiliar with double entry bookkeeping, the monetary version of the model of a trading economy would not be ideal. Then, bilateral balance might be the only means of insuring that individuals keep accurate records and balance their accounts and we would have to look for a minimum cost method of imposing bilateral balance. In such a situation, the principle would not change but the practice might well be to choose as a method of enforcing budget balance a commodity which is most portable, durable, divisible, and recognizable.

To the standard theory of value, the phenomenon of monetary exchange is surprising and distressing; surprising because the phenomenon is inexplicable and distressing because the phenomenon would seem to be one of the most elemental conclusions to be derived from any theory of exchange. Once we give up the standard theory framework which allows the execution of exchange to be the province of a centralized agency and concentrate on the logistics of more disaggregated trading arrangements, monetary exchange becomes explicable as a matter of course. It follows that these logistical considerations are worthy of attention by general equilibrium theorists.[29]

[26] See Radner.

[27] When individuals are groping for equilibrium exchange rates in a world where such voluntary restraint is lacking, monetary exchange would be essential. Imagine how much more difficult would be the approach to equilibrium if payment for one commodity were made in an arbitrary collection of other commodities. On this issue of price dynamics, I have benefited from reading Peter Howitt's study of stability in a decentralized regime of monetary exchange.

[28] See Brunner and Meltzer.

[29] See Donald Walker for a similar conclusion on the deficiencies of Walras' theory of exchange. Any reader familiar with the work of George Stigler and the contributors to Edmund Phelps et. al. will recognize that these logistical considerations have already received some attention as determinants of search unemployment in labor markets.

REFERENCES

K. J. Arrow and G. Debreu, "Existence of an Equilibrium for a Competitive Economy," *Econometrica*, July 1954, *22*, 265–90.

G. Bradley, "Trading Rules for a Decentralized Exchange Economy," rep. no. 55,

Yale Department of Administrative Sciences, May 1972.

K. Brunner and A. H. Meltzer, "The Uses of Money: Money in the Theory of an Exchange Economy," *Amer. Econ. Rev.*, Dec. 1971, *61*, 784–805.

R. W. Clower, "A Reconsideration of the Microfoundations of Monetary Theory," *Western Econ. J.*, Dec. 1967, *5*, 1–8.

———, "Theoretical Foundation of Monetary Policy," in G. Clayton et al., eds., *Monetary Theory and Monetary Policy in the 1970's*, London 1971, 15–28.

A. M. Feldman, "Non-Recontracting, Recontracting, and Equitable Trading Processes," unpublished doctoral dissertation, Johns Hopkins Univ. 1972.

F. H. Hahn, "On Some Problems of Proving the Existence of Equilibrium in a Money Economy," in Hahn and F. Brechling, eds., *The Theory of Interest Rates*, London 1965, 126–35.

———, "Equilibrium with Transactions Costs," *Econometrica*, May 1971, *39*, 417–40.

J. C. Harsanyi, "Games with Incomplete Information Played by Bayesian Players, I," *Manage. Sci.*, Nov. 1967, *14*, 159–82.

J. Hicks, "The Two Triads. Lecture I," in Hicks, ed., *Critical Essays in Monetary Theory*, London 1967, 1–16.

P. W. Howitt, "Stability and the Quantity Theory," unpublished, 1971.

M. Kurz, "Equilibrium with Transaction Cost and Money in a Single-Market Exchange Economy," tech. rep. no. 51, Institute of Mathematical Studies in the Social Sciences, Stanford Univ., Jan. 1972.

J. Marschak, "The Rationale of Money Demand and of 'Money Illusion'," *Metroeconomica*, Aug. 1950, *2*, 71–100.

J. F. Nash, "Non-cooperative Games," *Annals of Mathematics*, 1951, *54*, 155–62.

J. Niehans, "Money in a Static Theory of Optimal Payments Arrangements," *J. Money, Credit, Banking*, Nov. 1969, *1*, 796–26.

———, "Money and Barter in General Equilibrium with Transactions Costs," *Amer. Econ. Rev.*, Dec. 1971, *61*, 773–83.

J. M. Ostroy, "Exchange as an Economic Activity," unpublished doctoral dissertation, Northwestern Univ. 1970.

———, "The Informational Efficiency of Monetary Exchange," disc. pap. no. 15, UCLA Department of Economics, Nov. 1971.

——— and **R. M. Starr**, "Money and the Decentralization of Exchange," forthcoming.

D. Patinkin, "The Indeterminacy of Absolute Prices in Classical Economic Theory," *Econometrica*, Jan. 1949, *17*, 1–27.

E. S. Phelps et. al., *Microeconomic Foundations of Employment and Inflation Theory*, New York 1970.

R. Radner, "Competitive Equilibrium under Uncertainty," *Econometrica*, Jan. 1968, *38*, 31–58.

P. A. Samuelson, "The Pure Theory of Public Expenditure," *Rev. Econ. Statist.*, Nov. 1954, *36*, 387–89.

A. Smith, *The Wealth of Nations*, New York 1937.

K. C. Sontheimer, "On the Determination of Money Prices," *J. Money, Credit, Banking*, forthcoming.

R. M. Starr, "Equilibrium and the Demand for Media of Exchange in a Pure Exchange Economy with Transactions Costs," Cowles Foundation disc. pap. no. 300, Oct. 1970.

———, "Exchange in Barter and Money Economies," *Quart. J. Econ.*, May 1972, *86*, 290–302.

G. J. Stigler, "Information in the Labor Market," *J. Polit. Econ.*, Oct. 1962, *70*, 94–105.

E. C. H. Veendorp, "General Equilibrium Theory for a Barter Economy," *Western Econ. J.*, Mar. 1970, *8*, 1–23.

N. Wallace, "An Approach to the Study of Money and Nomoney Exchange Structures," disc. pap. no. 6, Center for Economic Research, Univ. of Minnesota, June 1971.

D. A. Walker, "Leon Walras in Light of His Correspondence and Related Papers," *J. Polit. Econ.*, July/Aug. 1970, *78*, 685–701.

L. Walras, *Elements d'Economie Politique Pure*, trans. W. Jaffe, *Elements of Pure Economics*, Homewood 1954.

Externalities, Liability, Separability, and Resource Allocation

By JAMES R. MARCHAND AND KEITH P. RUSSELL*

"... only through a more detailed understanding of the way that supply externalities affect allocative outcomes can the efficacy either of organizational alternatives such as centralized industry management or of corrective devices such as taxes be considered with sufficient sophistication to yield any hope for reasonably efficient results."

Charles J. Goetz and
James M. Buchanan, p. 883

The study of resource allocation in the presence of externalities continues to be of great interest to economists. Primarily because of the welfare implications which arise from models characterized by interdependencies in production and/or consumption, such models are of great importance to policy makers. The purpose of this paper is to analyze and compare the allocative properties of a model containing externalities in which there are a number of alternative environments facing decision makers. Specifically, we postulate the existence of a production externality and analyze the effects on resource allocation of a) the imposition (or possible imposition) of liability for such externalities; and b) the solicitation of compensation or a bribe from the affected party by the externality generating firm. These environments are analyzed for both separable and nonseparable cost functions.[1] Our results indicate that the allocation of resources in these alternative environments is directly related to the form of the cost function.

In a widely quoted paper, Ronald Coase provides an analysis of the resource allocation effects of the imposition of liability upon a situation characterized by externalities.[2] In particular, he argues that, in a two-firm externality situation, imposition of liability by a legal authority on *either* party results in an efficient allocation of resources (i.e., the allocation of resources is unaffected by the disposition of liability). This result seems to have been accepted without reservation, see for example, George Stigler, Guido Calabresi, and Thomas Borcherding. The results presented here purport to show that Coase's conclusions are more restrictive than is generally believed. Specifically, we show that Coase's conclusion regarding the neutrality of liability on resource allocation holds only in the restricted case of additively separable cost functions.[3]

The questions considered here, especially the crucial role of separability, would seem to have been alluded to but never completely analyzed in the previous literature. In their article in *Economica*, Buchanan and William Stubblebine anticipate the differences between separable

* Assistant professor of economics, California State University, Northridge, and research associate, Human Resources Research Center, University of Southern California, respectively. The comments of Vittorio Bonomo, James Buchanan, Daniel Orr, Steven Huxley, and an anonymous referee are gratefully acknowledged. We retain responsibility for remaining errors.

[1] A function $f(x_1, x_2)$ is said to be separable if it can be written as $f(x_1, x_2)=f^1(x_1)+f^2(x_2)$. See Abraham Charnes and Carlton Lemke and Otto Davis and Andrew Whinston (1962).

[2] See also James Buchanan (1966), Guido Calabresi, Charles Plott, and Ralph Turvey. Also, E. J. Mishan (1965, 1971) critically discusses Coase's work, though along different lines than those followed here.

[3] As with Coase's original article, this paper deals explicitly with the case of a relatively small number of firms. See Section III for a further discussion of this point.

and nonseparable functions, although their model was formulated in terms of utility functions. Translating their analysis into arguments in terms of production functions, the link between their paper and ours would lie in a simple theorem, which gives the relationship between production functions and cost separability.[4] In a recent piece, William Baumol argues in favor of Pigovian (unilateral) taxes and subsidies to achieve optimal resource allocation and, while not explicitly considering separability, provides a model which is somewhat analogous to one of ours (see fn. 10). Davis and Whinston (1962) discuss separable versus nonseparable cost functions and the practical feasibility of imposing taxes and subsidies to achieve optimality. Two recent papers, by Goetz and Buchanan and by Lawrence Schall, analyze the competitive equilibrium and Pareto optimality characteristics of supply externalities with models similar to those developed here (see also Yasusuke Murakami and Takashi Negishi). These works do not, however, explicitly explore how separability affects resource allocation under alternative liability arrangements, which is one of our main purposes here.

Our analysis proceeds as follows. We assume the existence of two different, but interacting, productive activities. First, traditional models of these productive activities under different decision units or firms are considered. Here "traditional" is taken to mean similar to (or identical to) those used explicitly or implicitly in previous literature such as Frank Knight's well-known article, "Some Fallacies in the Interpretation of Social Cost," or in any intermediate textbook discussion of external diseconomies between firms in different industries. Next, the same situation is analyzed using models with a more complete and expanded view of profit maximization, which have been suggested by Coase. Thirdly, the analysis returns to the profit maximizing framework of the first model, but the assessment of liability is changed from that implied in the first case. Finally, the analysis concludes with a model of the activities under joint production, that is, one firm. From the first-order conditions yielded by these models and with some further assumptions about the shape of cost curves, we obtain interesting implications about resource allocation to be compared and evaluated.

I. Resource Allocation in Alternative Environments: Nonseparable Costs

A. *Profit Maximization under Externalities*

Consider two firms, A and B, each a member of a different competitive industry, whose productive processes interact. Denote A and B outputs as q_1 and q_2 and their total cost functions as $A(q_1)$ and $B(q_1, q_2)$, respectively. (These functions relate the costs of the inputs purchased directly by the firms to the levels of output indicated in the arguments.) Assume that $A_1>0, B_1>0, B_2>0, A(0)=0$, and $B(q_1, 0)=0$.[5] That is, we assume that production by firm A will raise the costs of firm B at any strictly positive output level that B chooses to produce: a case of external diseconomies.[6]

[4] THEOREM: *The cost function $C(q_1, q_2)$ is separable i.e., $C(q_1, q_2)=C^1(q_1)+C^2(q_2)$ if, and only if, $q_1=F(x, q_2)$ is separable, i.e., if, and only if, $q_1=F^1(x)+F^2(q_2)$* where *x is an input vector for the production of q_1.*

The proof of this is based on necessary conditions for cost minimization and is omitted here.

[5] For notational convenience we will use subscripts on the functions to indicate *both* partial and total derivatives. That is, $A_1=dA/dq_1$ while $B_1=\partial B/\partial q_1$. The context will make clear if we are discussing partials or totals.

[6] The interaction here is one way, and interdependence is, of course, feasible. The form chosen here simplifies the mathematics without altering the conclusions. In addition, the question of the existence of an equilibrium (as discussed by Davis and Whinston (1962)) is avoided.

In the unrestricted situation firm B bears the cost of the interaction; A is not liable for the extra cost it inflicts on firm B. When A and B face the market-determined output prices P_1 and P_2, respectively, simple models of these firms are:

Firm A: $\underset{\{q_1\}}{\text{Maximize}}\ P_1q_1 - A(q_1)$

subject to $q_1 \geq 0$

Firm B: $\underset{\{q_2\}}{\text{Maximize}}\ P_2q_2 - B(q_1, q_2)$

Subject to $q_2 \geq 0$

Thus, conventional theory concludes that the firms' output decisions, $\hat{q}_1$ and $\hat{q}_2$, must be such that:

$$P_1 = A_1(\hat{q}_1) \tag{1}$$

$$P_2 = B_2(\hat{q}_1, \hat{q}_2) \tag{2}$$

Since $B_1(q_1, q_2) > 0$, when $\hat{q}_1 > 0$ the profit-maximizing output level for firm B is lower than it would be in the absence of the externality.[7] This is, of course, the standard result of traditional behavior in the presence of externalities.

B. *An Expanded View of Profit Maximization*

One of the main arguments of "The Problem of Social Cost" by Coase is that in the case of undefined liability, the above model of firm A is incorrect. Coase states that ". . . a receipt foregone of a given amount is the equivalent of a payment of the same amount" (p. 7), and that, furthermore, the "injured" firm, B, should be indifferent to whether it bears an extra cost or pays compensation of an equal amount to avoid this extra cost. Thus firm A's costs must include the bribe it is foregoing in choosing to produce a particular level of output, for this is an opportunity cost of production. Assuming that firm A would choose to produce $\hat{q}_1$, this cost is the compensation which B would be willing to pay and is equal to the cost saving B receives as a result of A's producing less than $\hat{q}_1$; that is, compensation would be $C = (B(\hat{q}_1, q_2) - B(q_1, q_2))$. This argument may be put another way in terms of the net revenue of the productive activities of firm A. By producing $\hat{q}_1$ units and selling them at a price of P_1 the firm obtains the gross revenue $P_1\hat{q}_1$. However, at the same time, if it does not solicit the bribe from firm B, it is foregoing revenue in the amount of $B(\hat{q}_1, q_2) - B(q_1, q_2)$ which is the amount B is willing to pay to be rid of the costs A is imposing by producing $\hat{q}_1$ instead of q_1. Thus, firm A's net revenue from producing $\hat{q}_1$ instead of producing q_1 is $P_1\hat{q}_1 - [B(\hat{q}_1, q_2) - B(q_1, q_2)]$. Taking either the cost or the revenue approach, the models for firms A and B become (assuming that transactions are costless):[8]

$$\underset{\{q_1\}}{\text{Maximize}}\ P_1q_1 - A(q_1) + [B(\hat{q}_1, q_2) - B(q_1, q_2)]$$

subject to $q_1 \geq 0$

$$\underset{\{q_2\}}{\text{Maximize}}\ P_2q_2 - B(q_1, q_2) - [B(\hat{q}_1, q_2) - B(q_1, q_2)]$$

subject to $q_2 \geq 0$

The output decisions of the firms, q_1'' and q_2'', are such that:

[7] We conduct our analyses in terms of explicit increases in the costs of production rather than decreases in the profit levels to facilitate the exposition of the role of separability in our model. Since profits are just revenues minus costs, this approach is perfectly general and involves no restrictions.

[8] The formulation given in the text was suggested by an anonymous referee. In our original formulation, A's criterion function included the opportunity bribe that he might attain over the various alternative quantities that he could produce, rather than the bribe received if A took $\hat{q}_1$ to be his initial output position. In this case, the model for firm A becomes: Maximize $P_1q_1 - A(q_1) - [B(q_1, q_2) - B(0, q_2)]$. With the model for B unchanged, the equilibrium conditions remain (3) and (4).

(3) $\quad P_1 = A_1(q_1'') + B_1(q_1'', q_2'')$

(4) $\quad P_2 = B_2(\hat{q}_1, q_2'')$

C. *The Imposition of Liability*

Let us now suppose that a law or a judge were to declare firm A liable for the extra costs experienced by firm B in a state of nature. Liability is therefore $L = B(q_1, q_2) - B(0, q_2)$. In these circumstances the appropriate models of the firms would be:[9]

Firm A: Maximize$_{\{q_1\}}$ $P_1q_1 - A(q_1) - [B(q_1, q_2) - B(0, q_2)]$

subject to $q_1 \geq 0$

Firm B: Maximize$_{\{q_2\}}$ $P_2q_2 - B(q_1, q_2) + [B(q_1, q_2) - B(0, q_2)] = P_2q_2 - B(0, q_2)$

subject to $q_2 \geq 0$

In this case it is necessary that for the chosen outputs, q_1' and q_2'

(5) $\quad P_1 = A_1(q_1') + B_1(q_1', q_2')$

(6) $\quad P_2 = B_2(0, q_2')$

D. *Resource Allocation under Centralized Decision Making*

To find the outputs which are socially optimal one has to find the outputs which maximize the social product. One way of doing this is to internalize the externality by making both activities subject to one decision unit. When both productive activities are subject to one decision unit, the damage done and the costs incurred in the Bth activity due to the Ath activity must be recognized when the decision maker chooses the output level for the Ath activity. Thus, a profit-maximizing model of the joint firm is:

Maximize$_{\{q_1, q_2\}}$ $P_1q_1 + P_2q_2 - A(q_1) - B(q_1, q_2)$

subject to $q_1 \geq 0$, $q_2 \geq 0$

For the socially optimal outputs, q_1^* and q_2^*, it is necessary that:

(7) $\quad P_1 = A_1(\overset{*}{q_1}) + B_1(\overset{*}{q_1}, \overset{*}{q_2})$

(8) $\quad P_2 = B_2(\overset{*}{q_1}, \overset{*}{q_2})$

There are thus four (possibly different) outcomes to the resource allocation process, corresponding to the four alternative environments. For purposes of comparison, in Table 1 we present the output levels corresponding to each environment.

TABLE 1

Environment	Output
Case a: Traditional Profit Maximization	$\hat{q}_1, \hat{q}_2$
Case b: Expanded Profit Maximization	q_1'', q_2''
Case c: Imposition of Liability	q_1', q_2'
Case d: Centralized Decision Making	q_1^*, q_2^*

Without making further assumptions about the second-order partial derivatives of the cost functions, little can be said about the relationship of q_1^*, q_2^*, the socially optimal outputs, to the levels of output chosen in the three cases cited. However, if we assume that the determinant of second-order partial derivatives

$$\begin{vmatrix} A_{11}(q_1) + B_{11}(q_1, q_2) & B_{21}(q_1, q_2) \\ B_{12}(q_1, q_2) & B_{22}(q_1, q_2) \end{vmatrix}$$

is nonzero for all $q_1, q_2 > 0$, then the transformations represented by the marginal cost functions $A_1(q_1) + B_1(q_1, q_2)$ and

[9] An alternative formulation of the problem is as follows: Let

$$\overline{B}(q_1, q_2) = \max_{\bar{q}_1} [B(\bar{q}_1, q_2) - B(q_1, q_2)]$$

$\overline{B}(q_1, q_2)$ is an upper limit for the bribe firm A could extract from firm B if joint production is (q_1, q_2). Then an alternative objective function for firm A is

$$\underset{\{q_1\}}{\text{Maximize}}\ P_1q_1 - A(q_1) + \overline{B}(q_1, q_2) = P_1q_1 - A(q_1) + \max_{\bar{q}_1} [B(\bar{q}_1, q_2) - B(q_1, q_2)]$$

This formulation leads to the same result as that given in the text.

$B_2(q_1, q_2)$ are one-to-one. With this property and the assumption that $A_{11}(q_1)$, $B_{11}(q_1, q_2)$, $B_{12}(q_1, q_2)$, $B_{21}(q_1, q_2)$ and $B_{22}(q_1, q_2)$ are positive, it can be shown (see Appendix) that conditions (9), (10), and (11) hold:

$$\hat{q}_1 > \overset{*}{q_1} \quad \text{and} \quad \hat{q}_2 < \overset{*}{q_2} \tag{9}$$

$$q_1' < \overset{*}{q_1} \quad \text{and} \quad q_2' > \overset{*}{q_2} \tag{10}$$

$$q_1'' > \overset{*}{q_1} \quad \text{and} \quad q_2'' < \overset{*}{q_2} \tag{11}$$

Conditions (9) indicate that for firms which individually profit maximize, the outputs forthcoming will not be socially optimal. In addition, in neither the compensation case nor in the liability case will the outputs be socially optimal. In the former case, firm A's output is too large (as in the first case) while firm B's output is too small. This nonoptimality is independent of whether firm A obtains the bribe: if he does not, we have condition (9); if he does, we have (11). In the liability case, and this is one of our most important results, examination of (5), (6), and (10) indicates the outcomes. In this situation, the equilibrium condition for firm A appears to be optimal, but analysis of the second firm's condition indicates that it is not socially correct. This is due to the fact that if firm B is fully compensated for any cost increase due to the actions of firm A, then B's privately optimal decision will be calculated as if firm A were producing zero output; q_2' will thus be socially too large and q_1' will be socially too small, the reverse of the compensation (bribe) case. This is in direct contradiction to what Coase concludes: in his paper, he denies that "... the fact that the cattle-raiser [firm A] would pay for all crops damaged would lead the farmer [firm B] to increase his planting" (p. 3). Notice also that there is a similarity between the compensation case and the liability case, but with opposite results. These order relationships are not surprising since they simply reflect the well-known tenet that output falls as either the marginal revenue curve shifts down or the marginal cost curve shifts up, all other things held equal.[10]

Thus, only in the case in which there is joint decision making is there an optimal allocation of resources. If legal forces make firm A totally liable, or if firm A either receives or does not receive the bribe, there is a misallocation of resources: in these cases the allocation of resources is nonoptimal. For the "state of nature" in which all of the costs of the externality accrue to firm B and A does not solicit the bribe, A's marginal cost is understated and B's is overstated, socially. The result is that at given market prices A overproduces and B underproduces in social terms. A similar result occurs if A receives compensation from B. Placing total liability on the firm "generating" the external diseconomies creates an inverse nonoptimal allocation of resources. Because firm B is being totally compensated, its marginal costs are understated in relation to social cost and firm B is induced to overproduce. As firm A is bearing total liability, its marginal costs are overstated and it produces less than the socially optimal amount of output.

[10] Baumol discusses the effects of *unilateral* taxes (subsidies) on resource allocation and shows that these taxes do in fact generate socially optimal allocations. He also discusses practical methods for implementing such taxes to achieve optimality. In our notation, his model would be similar to the liability case except for B's criterion function:

firm A: Maximize $P_1q_1 - A(q_1) - [B(q_1, q_2) - B(0, q_2)]$
firm B: Maximize $P_2q_2 - B(q_1, q_2)$

First-order conditions are

$$P_1 = A_1(\tilde{q}_1) + B_1(\tilde{q}_1, \tilde{q}_2)$$
$$P_2 = B_2(\tilde{q}_1, \tilde{q}_2)$$

which are identical to (7) and (8), i.e., the socially optimal allocations. Thus $\tilde{q}_1 = q^*$, $\tilde{q}_2 = q^*$. As shown below, however, with separable cost functions unilateral taxes are not the only means of generating a socially optimal allocation.

II. Resource Allocation in Alternative Environments: Separable Costs

With the framework developed in Section II, the analysis of the same types of environments but with separable cost functions is quite straight-forward.[11] Under the assumption of separability, firm B's cost function will now be written[12]

$$B(q_1, q_2) = b^1(q_1) + b^2(q_2) \text{ with } \overset{1}{b_1}, \overset{2}{b_1} \geq 0.$$

The remainder of the analysis consists simply of substituting this cost function into the previous models and determining the first-order necessary conditions. The models are as follows, with first-order conditions corresponding to (1)–(8) denoted by (1′)–(8′).

A. *Traditional Profit Maximization*

Under conditions of separability of the cost functions, models of the firms become:

Firm A: $\underset{\{q_1\}}{\text{Maximize}}\ P_1q_1 - A(q_1)$

Subject to $q_1 \geq 0$

Firm B: $\underset{\{q_2\}}{\text{Maximize}}\ P_2q_2 - b^1(q_1) - b^2(q_2)$

subject to $q_2 \geq 0$

In this case, the conditions corresponding to (1) and (2) above are:

$$(1') \qquad P_1 = A_1(\hat{q}_1)$$

$$(2') \qquad P_2 = \overset{2}{b_1}(\hat{q}_2)$$

B. *Compensation Case*

In this case, we have the following:

Firm A: $\underset{\{q_1\}}{\text{Maximize}}\ P_1q_1 - A(q_1) + [b^1(\hat{q}_1) + b^2(q_2) - b^1(q_1) - b^2(q_2)]$

subject to $q_1 \geq 0$

Firm B: $\underset{\{q_2\}}{\text{Maximize}}\ P_2q_2 - b^1(\hat{q}_1) - b^2(q_2)$

subject to $q_2 \geq 0$

Profit maximization conditions become

$$(3') \qquad P_1 = A_1(q_1'') + \overset{1}{b_1}(q_1'')$$

$$(4') \qquad P_2 = \overset{2}{b_1}(q_2'')$$

C. *The Imposition of Liability*

The models under conditions where liability is imposed are:

Firm A: $\underset{\{q_1\}}{\text{Maximize}}\ P_1q_1 - A(q_1) - [b^1(q_1) + b^2(q_2) - b^1(0) - b^2(q_2)]$

subject to $q_1 \geq 0$

Firm B: $\underset{\{q_2\}}{\text{Maximize}}\ P_2q_2 - b^1(0) - b^2(q_2)$

subject to $q_2 \geq 0$

Conditions (5) and (6) then become

$$(5') \qquad P_1 = A_1(q_1') + \overset{1}{b_1}(q_1')$$

$$(6') \qquad P_2 = \overset{2}{b_1}(q_2')$$

D. *Centralized Decision Making*

Finding conditions which maximize the social output give:

$\underset{\{q_1, q_2\}}{\text{Maximize}}\ P_1q_1 + P_2q_2 - A(q_1) - b^1(q_1) - b^2(q_2)$

subject to $q_1, q_2 \geq 0$

Maximum conditions are

$$(7') \qquad P_1 = A_1(\overset{*}{q_1}) + \overset{1}{b_1}(\overset{*}{q_1})$$

$$(8') \qquad P_2 = \overset{2}{b_1}(\overset{*}{q_2})$$

For this case, since $B_{12}(q_1, q_2) = B_{21}(q_1, q_2) = 0$, when one assumes that $A_{11}(q_1)$, $b^1_{11}(q_1)$, and $b^2_{11}(q_2)$ are positive for $q_1, q_2 > 0$,

[11] This is the case which Coase analyzed through the use of several numerical examples. His analysis concluded that the allocation of resources was independent of the definition of liability, but that this definition did determine the distribution of income (i.e., returns). See also Warren Nutter for additional numerical examples and an analysis similar to that of Coase.

[12] Recall that $b_1^1(q_1) = db^1/dq_1$; $b_1^2 = db_2^2/dq_2$.

it is possible to show that (see Appendix):

$$(9') \quad \overset{*}{q}_1 = q_1'' = q_1' \quad \text{and} \quad \overset{*}{q}_2 = q_2'' = q_2'$$

and

$$(10') \qquad \hat{q}_1 > \overset{*}{q}_1 \quad \text{and} \quad \hat{q}_2 = \overset{*}{q}_2$$

With separable cost functions the results of our analysis of resource allocation show lesser variation. If firm B's cost function is separable and firm A is the type of profit maximizer hypothesized by Coase, then the definition of liability for the costs of interaction does not affect the output decisions; they are socially optimal in either case. Only if firm A neither takes the bribe nor pays for the external costs it causes, is there a misallocation of resources.

III. Qualifications

It should be noted that the analysis which has been presented above has passed over several theoretical points and therefore must be qualified. The most serious shortcoming of this study is that it makes no distinction between *threatened* quantities and *actual* quantities.[13] In certain of the instances there may exist an incentive for firm A to threaten to produce a quantity greater than that which is optimal, in order to induce firm B to pay a higher bribe. (Of course, there are also cases in which firm A would have an incentive to understate his optimal output, for example, in the case of assigned liabilities.) In certain instances the [heroic] assumptions of perfect knowledge might vitiate this criticism. The analysis of a world of uncertainty where we can determine bounds on threatened quantities and determine each firm's optimal strategy using game-theoretic methods will be the topic of a forthcoming paper.

[13] On this, see G. A. Mumey. His conclusions that misallocations may arise from threats used as revenue generators and from coercive income distributions are independent of our conclusions.

Another point is that it may be quite fictitious to analyze two competitive firms in this type of situation. One either must leave the two-firm framework or drop the assumption of competitive organization. The homogeneity postulate of the theory of perfect competition requires that all firms in an industry face the same total cost situation. Thus, it is only theoretically correct to anlyze the interaction of two competitive industries. This extension would yield the same qualitative results for the individual firm. However, changing the firms' marginal cost curves requires changing the industries' supply curves implying that the price of the output and the number of firms would differ with different assignments of liability. This, of course, only further magnifies the shifts in the allocation of resources.

If one chooses to remain in a two-firm framework, then one might examine models of monopoly. This problem, however, seems to have been adequately analyzed in the literature (see, for example, Buchanan (1969)).

Also, though both firms were assumed to be members of competitive industries, the cost functions used here must be inclusive of Ricardian rents. This is necessary because, as Stanislaw Wellisz has said:

> In a perfectly competitive industry in which there are no Ricardian rents, external diseconomies result in factor use changes without creating any social costs. In such a setting every firm operates at a zero profit level and all the factors are paid their opportunity costs. If external costs are imposed upon one of the competitive units, that unit will make losses and in the long run go out of business. [p. 350]

Finally, it is possible that one might argue against our results on the grounds that we did not allow the individuals concerned sufficient scope to negotiate. That is, no matter what liability rule is in effect, any arrangements are preliminary in

nature and lead to more complex negotiations such that it is possible for individuals to merge and attain the jointly optimal solution. This notion of gains-from-trade in externality situations has been discussed by Buchanan and Stubblebine and by Baumol, and it is indeed possible that within the context of such negotiations a "generalized" Coase theorem might hold: that is, the jointly optimal solution is obtained independently of the assignment of liability or the liability rule in effect. This result would apply to both separable and nonseparable cases.

There are, however, several points to be considered before one can assume the validity of the generalized Coase theorem. In the first place, as noted by Baumol, in any practical situation the large numbers of transactors involved may make negotiation prohibitively expensive. In this case, the burden of proof in favor of a generalized Coase theorem would seem to lie with those who can show how such bargaining could take place: one cannot just assume that it does. Secondly, as noted in Baumol's paper, once such (private) gains are realized the resulting solution may in fact be nonoptimal (though, as Baumol shows, not through income effects as Buchanan and Stubblebine originally implied). Thus, while not rejecting the possibility of a generalized Coase theorem, we recognize that such a theorem could exist only under more fully specified conditions.

IV. Conclusion

The analysis above was begun with the idea of clarifying some issues concerning the type of resource allocation obtained under technological externalities in production. In particular, our analysis clarifies the question of the effect of making externality "generators" liable for external diseconomies. Neither from Coase's original analysis nor from the analysis developed herein can one universally say as Coase has said:

> It is necessary to know whether the damaging business is liable or not for damage caused since without the establishment of this initial delineation of rights there can be no market transactions to transfer and recombine them. But the ultimate result [i.e., the allocation of resources] (which maximizes the value of production) is independent of the legal position if the pricing system is assumed to work without cost. [p. 8]

This type of statement can only be made in the simple framework of separable cost functions. In the more complicated, yet more realistic world where production externalities involve nonseparable cost functions, the allocation of resources is not independent of the legal framework. As the inequalities in (9)–(11) show, the allocation of resources depends on the definition of liability.

In addition to indicating conditions under which Coase's conclusion does and does not hold, the analysis of the nonseparable cases points to the following summary The problem of social costs is a problem of joint costs. As Coase has said, "If we are to discuss the problem in terms of causation, both parties should take the harmful effect (the nuisance) into account in deciding on their courses of action" (p. 13). Firms on both sides of the problem must adjust their factor inputs and outputs to get the proper resource allocation. If one or the other side of the problem bears these costs alone, there will be a misallocation of resources.

The growth of problems involving questions of externalities, liability, and decision making under alternative institutional structures and the continuing interest of economists, policy makers, and the general public makes it imperative that the economics of these issues be made clear.

Appendix

In this Appendix we present simple proofs which are used in the body of the paper to indicate the relationships between output levels under alternative regimes.

Proof of (9)

1. For a given P_1, P_2 we can equate the right sides of equations (1) and (7) yielding

$$A_1(\hat{q}_1) = A_1(\overset{*}{q_1}) + B_1(\overset{*}{q_1}, \overset{*}{q_2})$$

Since $B_1(q_1^*, q_2^*) > 0$ by assumption, then $A_1(\hat{q}_1) > A_1(q_1^*)$. Assuming that $A_{11}(q_1) > 0$ then, $\hat{q}_1 > q_1^*$.

2. Equating the right half of equations (2) and (8) we find $B_1(\hat{q}_1, \hat{q}_2) = B_2(q_1^*, q_2^*)$. Since by assumption $B_{21}(q_1, q_2) > 0$ and $B_{22}(q_1, q_2) > 0$, then $\hat{q}_1 > q_1^*$ implies that $q_2^* < \hat{q}_2$.

Proof of (10)

1. By equating the right half of (6) and (8) we have that $B_2(0, q_2') = B_2(q_1^*, q_2^*)$. Since $q_1^* > 0$ and by assumption $B_{21}(q_1, q_2) > 0$ and $B_{22}(q_1, q_2) > 0$, then $q_2' > q_2^*$.

2. Equating the right half of (5) and (7) one finds

$$A_1(q_1') + B_1(q_1', q_2') = A_1(\overset{*}{q_1}) + B_1(\overset{*}{q_1}, \overset{*}{q_2})$$

Since we have assumed $B_{12}(q_1, q_2) > 0$ and $q_2' > q_2^*$, $B_1(q_1^*, q_2') > B_1(q_1^*, q_2^*)$. With $B_{11}(q_1, q_2) > 0$ suppose $q_1' \geq q_1^*$, then $B_1(q_1', q_2') \geq B_1(q_1^*, q_2') > B_1(q_1^*, q_2^*)$. Now $q_1' \geq q_1^*$ implies since $A_{11}(q_1) > 0$, that $A_1(q_1') \geq A_1(q_1^*)$. Subtracting this inequality from the first equation of this section one finds that $B_1(q_2', q_2') \leq B_1(q_1^*, q_2^*)$. This is a contradiction, thus we must have that $q_1' < q_1^*$.

Proof of (11)

1. Equating the right-hand sides of equations (4) and (8), we have $B_2(\hat{q}_1, q_2'') = B_2(q_1^*, q_2^*)$. Since $\hat{q}_1 > q_1^*$ and $B_{21}(q_1, q_2) > 0$, then $q_2'' < q_2^*$

2. Using equations (3) and (7), we have

$$A_1(q_1'') + B_1(q_1'', q_2'') = A_1(\overset{*}{q_1}) + B_1(\overset{*}{q_1}, \overset{*}{q_2})$$

The condition that $q_2'' < q_2^*$ requires for equality that $q_1'' > q_1^*$.

Proof of (9′)

The assumption of positivity of all the second-order partials is sufficient to insure that the Jacobian of the marginal cost functions is nonzero. This means that the two marginal cost functions represent a one-to-one transformation and that for the point P_1, P_2 in the image the arguments of the marginal cost functions must be unique. Thus

$$q_1'' = \overset{*}{q_1} = q_1' \quad \text{and} \quad q_2'' = \overset{*}{q_2} = q_2'$$

Proof of (10′)

1. Since our basic assumption is that $b_{11}^2(q_2) > 0$, $b_1^2(q_2)$ is a monotonic function. Thus $\hat{q}_2 = q_2^*$.

2. Through equating the right side of (1′) and (7′) one determines that $A_1(\hat{q}_1) = A_1(q_1^*) + b_1^1(q_1^*)$. By assumption $b_1^1(q_1^*) > 0$, thus $A_1(\hat{q}_1) > A_1(q_1^*)$. Since $A_{11}(q_1) > 0$ for $q_1 > 0$, then $\hat{q}_1 > q_1^*$

REFERENCES

W. Baumol, "On Taxation and the Control of Externalities," *Amer. Econ. Rev.*, June 1972, *62*, 307–22.

T. Borcherding, "Liability in Law and Economics: Note," *Amer. Econ. Rev.*, Dec. 1970, *60*, 946–48.

J. M. Buchanan, "External Diseconomies, Corrective Taxes, and Market Structure," *Amer. Econ. Rev.*, Mar. 1969, *59*, 174–77.

———, "Joint Supply, Externalities, and Optimality," *Economica*, Nov. 1966, *33*, 404–15.

——— and W. C. Stubblebine, "Externality," *Economica*, Nov. 1962, *29*, 371–84.

G. Calabresi, "Transaction Costs, Resource Allocation, and Liability Rules," *J. Law Econ.*, Apr. 1968, *11*, 67–73.

A. Charnes and C. E. Lemke, "Minimization of Non-Linear Separable Convex Functionals," *Nav. Res. Log. Quart.*, Dec. 1954, *2*, 301–12.

R. Coase, "The Problem of Social Cost," *J. Law Econ.*, Oct. 1960, *3*, 1–44.

O. Davis and A. Whinston, "Externalities, Welfare, and the Theory of Games," *J. Polit. Econ.*, June 1962, *70*, 241–62.

——— and ———, "On Externalities, Information and the Government-Assisted Invisible Hand," *Economica*, Aug. 1966, *33*, 303–18.

C. J. Goetz and J. M. Buchanan, "External Diseconomies in Competitive Supply," *Amer. Econ. Rev.*, Dec. 1971, *61*, 883–90.

F. Knight, "Some Fallacies in the Interpretation of Social Cost," *Quart. J. Econ.*, Aug. 1924, *38*, 582–606.

E. J. Mishan, "Reflections on Recent Developments in the Concept of External Effects," *Can. J. Econ.*, Feb. 1965, *31*, 3–34.

———, "The Postwar Literature on Externalities: An Interpretative Essay," *J. Econ. Lit.*, Mar. 1971, *14*, 1–29.

G. A. Mumey, "The 'Coase Theorem': A Reexamination," *Quart. J. Econ.*, Nov. 1971, *85*, 718–23.

Y. Murakami and T. Negishi, "A Note on a Formulation of External Economy," *Int. Econ. Rev.*, Sept. 1964, *5*, 328–34.

G. W. Nutter, "The Coase Theorem on Social Cost: A Footnote," *J. Law Econ.*, Oct., 1968, *11*, 503–07.

C. R. Plott, "Externalities and Corrective Taxes," *Economica*, Feb. 1966, *33*, 84–87.

L. D. Schall, "Technological Externalities and Resource Allocation," *J. Polit. Econ.*, Sept./Oct. 1971, *79*, 983–1001.

G. Stigler, *The Theory of Price*, 3d ed., New York 1970.

R. Turvey, "On Divergences between Social Cost and Private Cost," *Economica*, Aug. 1963, *30*, 309–13.

S. Wellisz, "On External Diseconomies and the Government-Assisted Invisible Hand," *Economica*, Nov. 1964, *31*, 345–62.

Assets, Subsistence, and The Supply Curve of Labor

By Yoram Barzel and Richard J. McDonald*

The "backward bending" supply curve of labor is now accepted as a matter of course by most economists. It has no doubt been perplexing to observe that the most commonly employed types of utility functions do not yield such curves under the usual textbook analysis of the problem.[1] Particular preference maps have been found that generate backward bending curves;[2] however, they are nonparametric, leading to difficulties of estimation, and upon closer examination seem to imply counter-intuitive results. We will show that taking into account the wealth position of an individual on the one hand and survival consideration on the other greatly expands the variety of shapes that can be derived for the supply curve from some simple utility functions. The use of a specific simple utility function also implies some severe restrictions on the form the supply curve can take, rendering it testable. Empirical evidence is shown to support the conclusion that the supply curve is monotonic. We will also show that the notion that the aggregate supply curve of labor slopes down rests, in part, on an error of aggregation, and that the empirical evidence usually cited in support of the negative slope, when correctly interpreted, cannot be so construed.

I. The Supply Curve of the Individual

A curious method is commonly employed to derive the accepted shape of the supply curve of labor. The typical text first points out (following Lionel Robbins) that a wage change results in an income as well as a substitution effect and, noting the importance of leisure in the individual budget, concludes reasonably enough that the supply curve may, in some of its range, have a negative slope. But then the discussion proceeds without further analysis to suggest something about a turning point in the curve—that the curve will turn back only after an initial positive slope. In Milton Friedman's words, " . . *beyond* some point the income effect dominates the substitution effect" (1966, p. 204, italics added) and the change in sign is explained by the statement that ". . . in a primitive society, the initial low wage rate at which the income effect becomes dominant reflects a lack of familiarity with market goods and a limited range of tastes. As tastes develop and knowledge spreads, the point at which the income effect dominates tends to rise." The sign change seems to apply only to a primitive society, the value at which this occurs seems to shift around, and its explanation is rather lame. Although Friedman is only one of many economists to accept uncritically the backward bending shape of the supply curve,[3] his argument is singled out precisely because of his usual astuteness and the advanced nature of the text.[4]

* University of Washington. Some of the work on this paper was done while Barzel was visiting University College, London, supported by a Ford Foundation Fellowship.

[1] For example, a Cobb-Douglas utility function yields a perfectly inelastic supply curve.

[2] See, for example, Giora Hanoch.

[3] J. R. Hicks apparently was the first to introduce the backward bend, but he failed to offer any satisfactory explanation.

[4] To cite one more example, Paul Samuelson in his elementary text also subscribes to the backward bending shape of the supply curve.

Consider now a point on the supply curve of labor and observe what happens to the income change due to *successive* wage increases. The effect of the first wage increase is independent of whether the supply curve at that point has a positive or a negative slope, not those of the second and subsequent wage changes. If the supply curve has a positive slope, the number of hours affected by subsequent wage changes is larger and, other things equal, the income effect tends to grow stronger and stronger.[5] The converse will occur if the income effect dominates in the first place, rendering the slope of the supply curve negative. These "self-correcting" tendencies give some basis for expecting the supply curve to eventually become vertical.

The asset holdings of the individual play an important (and thus far largely neglected) role in determining the shape of the supply curve of labor.[6] If we consider an extremely wealthy individual, it seems intuitively clear that as long as work itself is not a commodity, no labor will be offered at the lowest range of wages; assuming continuity of preferences and the absence of indivisibilities, an infinitesimal amount will be supplied at the point of entry into the labor market. From there, at least for awhile, the curve *has* to slope upwards.

On the other hand, for an individual with no wealth whatsoever and no income source other than his own work, the very lowest wage will be insufficient for survival. As wages increase, a point will be reached where survival becomes possible if he supplies the highest physically possible amount of labor. For such an individual, as the wage rate continues to rise, the amount of labor supplied *cannot* increase; given that leisure is a commodity, the supply curve has to have a negative slope right from its *very beginning*. It will be suggested in the following analysis that we predominantly observe just such an initially negatively sloped curve eventually tending to become perfectly inelastic.

To proceed with the formal analysis, we consider an individual who derives satisfaction from the consumption of two goods: market purchased commodities, denoted by C, and leisure, denoted by R. Assume that the preferences of the individual can be characterized by a function $f(C', R')$, which at a point (C', R') in the commodity space indicates the individual's marginal rate of substitution between the two commodities at that point.

That is,

$$\frac{dC'}{dR'} = -f(C', R')$$

for small movements leaving the consumer as well off as he was before. The accepted range of this function is for $C', R' \geq 0$. But it should be recognized that unless some positive level of consumption of C is reached, survival is not possible, and so for some positive values of C a preference map cannot be said to exist. Similarly, survival considerations may dictate a certain minimal level of leisure (or rest) time. Notice that the roles played by the two survival requirements are not symmetric since all individuals are endowed with more time than is needed for survival but not all have sufficient assets for survival. Denote by S the minimal required consumption of goods per day, and by T the minimal required leisure time and define C and R as $C = C' - S$ and $R = R' - T$. We now assume that the arguments in the marginal rate of substitution are C and R so that

[5] Of course, it is possible that as wages rise the rate of change of labor supply, with respect to a change in income, may change sufficiently to negate this tendency.

[6] Kenneth Boulding notes the role of assets but does not proceed to examine it fully (pp. 800–01). He, too, draws a backward bending supply curve of labor even though his illustrations (pp. 210–11) demonstrate only a negative slope and not a turning point.

$$(1) \qquad \frac{dC}{dR} = -f(C, R)$$

along an indifference curve, for C, $R>0$. This transformation, while innocuous as long as f is not further specified, is a substantive one once specific properties are assumed for the preference map, as we shall do below. We also assume that $f(C, R)$ is positive, differentiable, and that preferences are characterized by a diminishing marginal rate of substitution. Thus

$$(2) \quad \frac{d^2C}{dR^2} = f(C, R)f_1(C, R) - f_2(C, R) > 0$$

for movements along an indifference curve, and for all C, $R>0$, where f_1 and f_2 are, respectively, the partial derivatives of f with respect to C and R.

The individual is subject to constraints on time and expenditures. The time constraint is $R+T+L=D=24$, where D is the length of the day and L is labor hours.[7] Given the notion of required rest time, it is more convenient to write the constraint as

$$(3) \qquad R + L = D - T = D'$$

where D' is the fixed number of hours whose composition can be allocated to leisure or labor.

The expenditure constraint is $Y=WL+PA=PC'$, where Y is money income, W is the money wage rate (or the shadow wage if the individual is self-employed), P is the price of market commodities, and A is per day nonwage income in units of consumption goods. By rearranging and substituting from (3) for L, we get

$$(4) \qquad C + wR = wD' + A - S$$

where $w=W/P$. Equation (4) has as its variables C and R, the two arguments in the marginal rate of substitution. Notice that if $wD'<S-A$, then even if the individual works the physically maximum number of hours, he will not earn enough to survive. In other words, unless the condition

$$(5) \qquad w > \frac{S - A}{D'}$$

is satisfied, C and R cannot both be positive, and the function f is not defined.

We assume that the individual chooses the most preferred combination that he can attain, given that he is able to survive.[8] Necessary and sufficient conditions for this to be the bundle (C_0, R_0), given (5), are that C_0 and R_0 satisfy the equations (2) and (4), together with

$$(6) \qquad f(C_0, R_0) = w$$

This last condition is independent of the transformation previously imposed on the origin of the function f.

Comparative statics results can be obtained by differentiating conditions (4) and (6) partially, first with respect to A and then with respect to w. The resulting slopes of the demands for consumption and leisure with respect to the wage rate are

$$(7) \quad \frac{\partial C}{\partial w} = f/(ff_1 - f_2) + (D' - R)\frac{\partial C}{\partial A}$$

$$(8) \quad \frac{\partial R}{\partial w} = -1/(ff_1 - f_2) + (D' - R)\frac{\partial R}{\partial A}$$

Now we want to concentrate on the effects of wage changes on the quantity of labor supplied. Because of equation (3) we can write

$$(9) \quad \frac{\partial L}{\partial w} = 1/(ff_1 - f_2) - (D' - R)\frac{\partial R}{\partial A}$$

The first term of the right-hand side of

[7] No substantive change is involved if we carry the analysis in terms of some other time unit, such as a week or a year. More importantly, our analysis is equally relevant to market-employed as to self-employed individuals.

[8] If $wD'=(S-A)$, the consumer will thus choose $C=0$, $L=D'$ rather than to work less and starve.

(9) is the pure substitution effect and is necessarily positive, due to the convexity of the preference map. The second term, assuming that leisure is not an inferior commodity, is also positive but is preceded by a minus sign. It is clear from this formulation that without further specification of the consumer's preferences, it is impossible to say which of the two effects will dominate.[9] Notice, however, that if we are in a region where the income effect dominates, as w increases the term $(D'-R')$ is declining, tending to diminish the strength of the entire income term. The converse is true if the substitution effect dominates.

We now introduce a more specific form of the consumer's preference. This will allow more definite and more readily refutable empirical implications. The more detailed specifications will be introduced in two steps. First, the function $f(C, R)$ will be restricted to be homothetic. In this case the marginal rate of substitution between (net) consumption and leisure is a function only of the ratio in which the two goods are consumed. Equation (9) can now be rewritten as

$$\frac{\partial L}{\partial w}=\frac{1}{w}\,\frac{R}{(C+wR)}\,[C(\sigma-1)+(A-S)] \tag{10}$$

where σ is the elasticity of substitution between net consumption and net leisure.[10] So the slope of the supply curve depends on the signs of $(\sigma-1)$ and of $(A-S)$.[11] In the special case $\sigma=1$, where the preferences imply a Cobb-Douglas type of utility function, the sign of the slope of the supply curve hinges entirely on the sign of $(A-S)$. In general σ need not be constant but will vary with the ratio C/R. Letting $x=C/R$, it can be shown that

$$\frac{\partial\sigma}{\partial w}=\frac{1}{x}\,(1-\sigma-\sigma\epsilon)$$

where ϵ (itself a function of x) is defined by

$$\epsilon=\frac{h''(x)\cdot x}{h'(x)}$$

and $h(x)$ is the marginal rate of substitution as defined in the Appendix. Then, using the definition of σ, we have

$$\frac{\partial\sigma}{\partial w}=\frac{\sigma}{w}\,(1-\sigma-\sigma\epsilon)$$

and since our assumptions about the preference map impose no restrictions on the sign or magnitude of ϵ, nothing a priori can be said about the direction of change of σ as wages vary. Thus, if we do not restrict σ to a constant value, the supply curve can assume virtually any shape and therefore will have no testable implications.

The second step in specifying preferences, then, is to constrain σ to a constant value. This will also further simplify matters and allow for easy graphical interpretation. The marginal rate of substitution takes the particular form

$$f(C, R)=\left(\frac{1-\alpha}{\alpha}\right)\left(\frac{C}{R}\right)^{1/\sigma} \tag{11}$$

where $0<\alpha<1$.[12] The above assumptions

[9] If consumption C is assumed to be a normal good, then the demand for it (with respect to w) must be positively sloped, due to the restriction that the individual can consume no more than D' hours of leisure.

[10] See the Appendix.

[11] The terms in the square brackets in (10) are grouped differently than is conventional. Since $A-S-C=-wL$, these terms represent the income effect, while $C\sigma$ represents the substitution effect.

[12] Equation (11) would follow from any monotonically increasing function of the *CES* utility function.

$$V(C, R)=[\alpha C^{-\beta}+(1-\alpha)R^{-\beta}]^{-1/\beta} \qquad \text{for } \beta\neq 0$$
$$V(C, R)=C^{\alpha}R^{1-\alpha} \qquad \text{for } \beta=0$$

defined for $0<\alpha<1$ and $-1<\beta<\infty$. The elasticity of substitution is then equal to $1/(1+\beta)$. This specification is a generalization of the class of utility functions generating the so-called linear expenditure system of demand equations introduced by Laurence Klein and Herman Rubin, and more recently adopted by Richard Stone.

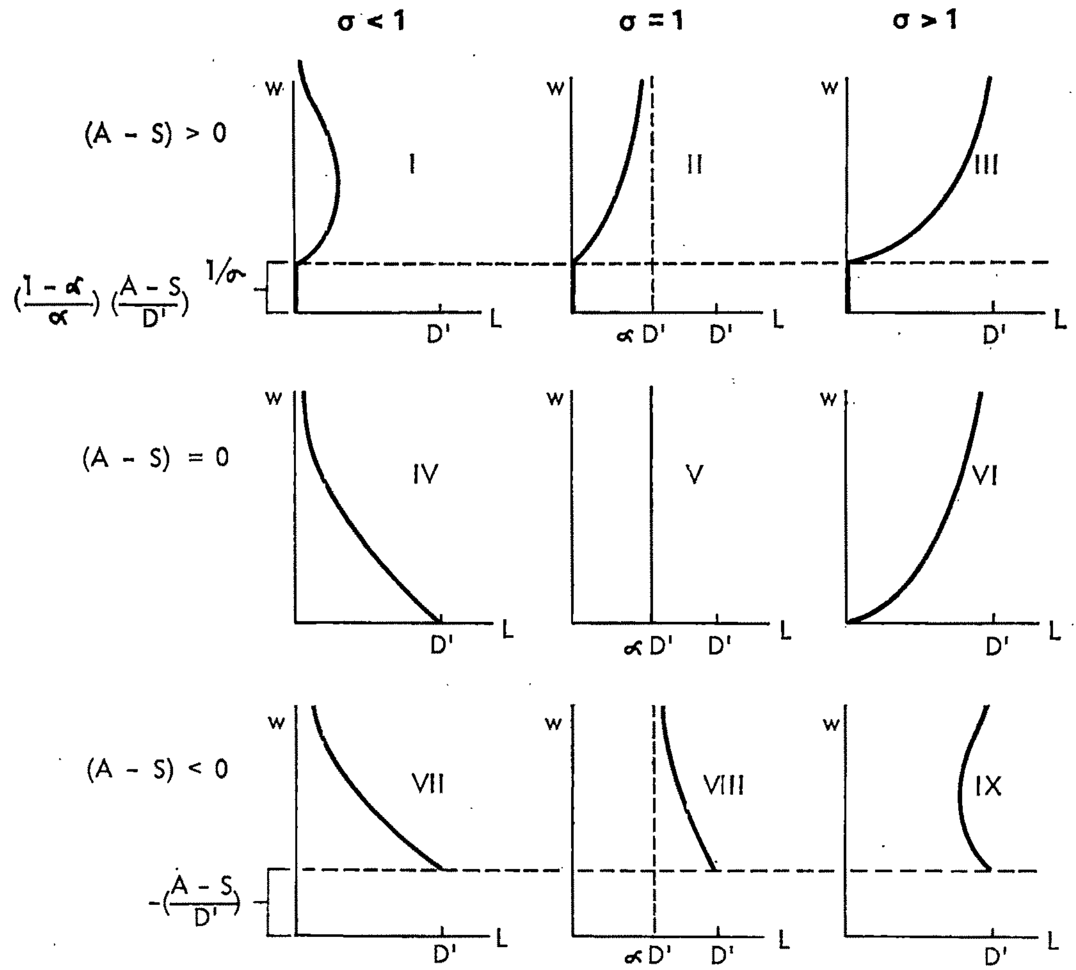


FIGURE 1

may appear to be highly restrictive; nevertheless, depending on the signs of $(\sigma-1)$ and $(A-S)$, we get the textbook characterizations of the supply curve as well as some novel ones.

The nine panels in Figure 1 are drawn for cases where $\sigma \gtreqless 1$ in combination with $(A-S) \gtreqless 0$. This set of diagrams gives us a whole gamut of possible results—monotonically rising and falling, perfectly inelastic, and both backward and forward bending. So the assumptions made regarding the form of the preference map do not turn out to be as restrictive as might first appear.[13] Notice that in the three figures with $(A-S)<0$ the supply curve is not defined for wage rates below $(S-A)/D'$.

The second row of panels (IV–VI) where $A-S=0$ corresponds to the conventional treatment of the supply of labor where both A and S are implicitly assumed equal to zero. It is easy to see, by comparing the second row with the first and with the third, the crucial importance of these two

[13] One important restriction is the linearity of the Engel curves. See Samuelson (1948). Given our subsistence requirements, the Engel curves do not necessarily go through the origin.

factors.[14] If assets are large enough, the supply curve will always have an initial region with a positive slope (and will be monotonic for $\sigma \geq 1$). If assets alone are insufficient to provide for survival, the supply curve will always slope negatively initially (and will be monotonic for $\sigma \leq 1$).[15] Consider an individual possessing no income source other than his own labor. If we do not explicitly take into account survival considerations, then the second row will appear pertinent. An individual with a utility function leading to $\sigma > 1$ would, at the lower wage range, then prefer to die from starvation rather than to earn enough for survival. This absurd result points to the reasonableness of transforming the preference function as we did.

There is no a priori reason to accept any one of these cases. It is important to note, however, that since few individuals, even in Western countries, start their working career with asset holdings sufficient for survival, the third row of diagrams represents the wealth conditions of most individuals. For some individuals, however, $A > S$; for those, we should observe, at least in part of the range, a positive relation between wages and hours, and in such cases the vertical intercept itself will rise with the level of asset holdings. It will be shown that empirical evidence conforms closely to the *C-D* case illustrated in Panel VIII.

II. The Negative Income Tax and the Supply of Labor

The preceding analysis can easily be adapted to determine the effects on the supply of labor of a variety of tax and transfer schemes. Of current interest are those tagged as "guaranteed annual income" or "negative income tax." Under most of these the individual whose income exceeds some "break-even" level, denoted here by Y_b, is being taxed and the individual whose income falls short of that level is being subsidized. For simplicity let us assume that the marginal rate of tax does not vary with income, that is, the rate of tax (which is also the rate of subsidy) is constant. If we denote this rate by t, the floor on income is tY_b.

The disposable income available to the individual, denoted by Y_d, is given then by

$$Y_d = Y + t(Y_b - Y) \tag{12}$$

where $Y = wL + A$. The budget constraint now becomes, after manipulations similar to those used to derive (4),

$$C + (1 - t)wR = (1 - t)(wD' + A) + tY_b - S \tag{13}$$

$$= (1 - t)wD' + (A - S) + t(Y_b - A) \tag{13'}$$

The condition for an interior maximum is now

$$f(C, R) = (1 - t)w \tag{14}$$

Again, assume the preference map to be homothetic, then the same comparative static techniques used to derive (10) give us

$$\frac{\partial L}{\partial w} = \frac{1}{w} \frac{R}{[C + (1 - t)wR]} \{C(\sigma - 1) + [(1 - t)A - (S - tY_b)]\} \tag{15}$$

This result is essentially the same as (10), with the wage rate deflated by the tax (and subsidy) rate, and with the term in the square brackets corresponding to $[A - S]$ in (10). The term $(1-t)A$ is net-of-tax nonwage income, and $(S - tY_B)$ is the portion of subsistence not covered by

[14] In analyzing the labor-leisure problem many authors implicitly assume that labor is the only source of income. As can be seen by comparing the second with the first (and third) row of diagrams, this may result in a serious error.

[15] An important omission in our analysis is the possibility that an individual may be able to forestall starvation by consuming part of his stock of wealth in addition to the flow of income resulting from it. This is, of course, only a short-run solution. A fuller treatment of this problem would require bringing time considerations explicitly into the analysis.

the guaranteed minimum income. As before, the sign of the term in the outer brackets determines the sign of $\partial L/\partial w$. For the case where $\sigma=1$, the supply curve is positively sloped if net nonlabor income exceeds the "net" subsistence term. It is not possible to determine a priori the effect of the tax plan on this final term, since the tax lowers net nonlabor income while the existence of the income floor weakens the force of the survival requirements.

Given, however, the notion of a minimum survival requirement, it would seem that if a guaranteed income has any meaning at all, it has to at least suffice for survival. We want, then, to examine the case where $tY_b \geq S$. The terms in square brackets in equation (15) must then be positive. The possible cases then reduce to those in the first row of Figure 1 (with trivial modifications in interpretation of the terms). The supply curve will always slope upwards initially, and may slope upwards throughout. Only when σ is sufficiently less than one is it possible for the labor supply to turn into a negatively sloping curve.

To understand the effects of the plan more completely, it is helpful to find the effect of a change in the tax rate on the supply of labor. This can be done by differentiating equations (13) and (14) partially with respect to the tax rate t. The result of these operations is shown in equation (16).

$$\frac{\partial L}{\partial t} = \frac{R[C(1-\sigma)+(S-Y_b)]}{(1-t)[C+(1-t)wR]} \tag{16}$$

Again the effect on the labor supply cannot be determined a priori, and it would take another series of nine figures to depict all of the possibilities. Notice that it is the break-even level of income, and not the floor (tY_b) on income, that is opposed to survival consumption S. Since the break-even level Y_b must be greater than the floor tY_b, and the floor in turn must exceed subsistence (see preceding paragraph), it is plausible that the break-even level will exceed subsistence. If this is the case, the effect of an increase in the tax (and subsidy) rate will be to decrease the labor supply (unless σ is sufficiently smaller than unity). Note also that at $Y=Y_b$ the tax is zero; consequently, for $Y=Y_b$ a change in the tax rate leads to a substitution effect but not to an income effect. As a final result, the partial effect of changes in the level of break-even income on labor supply can be found to be

$$\frac{\partial L}{\partial Y_b} = \frac{-tR}{[C+(1-t)wR]} \tag{17}$$

which is negative, given that $0<t<1$.

III. The Aggregate Supply Curve

The individual's demand for leisure is derived under the assumption that (in competition) the wage rate is an exogenously determined parameter; a change in the wage rate will lead to an income or wealth effect, in addition to the substitution effect. The central interest in this paper is with the uncompensated demand curve. So the demand for leisure in (8) and subsequently contains both a substitution and an income effect. It is well known that when considering the aggregate demand for a commodity, price is an endogenously determined parameter and, more importantly, that the net wealth effect will tend to zero. This recognition is one of the major contributions of general equilibrium analysis. Milton Friedman (1949) and subsequently Martin Bailey, in identically titled articles, "The Marshallian Demand Curve," both stress that in the aggregate, at a point in time, wealth effects tend to wash out due to the constraint imposed by the resource endowment of the economy. And in the absence of wealth effects, no Giffen goods are possible, and the aggregate demand curve for leisure, as for any other good, has to

slope down throughout.[16] But elsewhere in their writings, when drawing the supply curve of labor, both Friedman and Bailey allow for a negatively sloped portion of the curve. Friedman (1966) does not specify the level of aggregation, but Bailey (1962) clearly discusses the economy-wide supply curve of labor.

A possible explanation for such inconsistency is that empirical evidence strongly indicates a negative relation between the wage rate and the supply of labor. To the extent that this relation is observed in cross-sections, it is usually obtained for subgroups in truly partial equilibrium settings, and the evidence is not pertinent in determining the shape of the aggregate relation. As for time-series, the *ceteris paribus* is clearly violated. Comparisons are made between points in time for which the total productive capacity of the economy is not held constant. In other words, over time we observe not only the substitution due to an increase in real wages but also the effects of a net increase in real wealth. The wealth increase is not the result of the change in wage; if anything, it is the cause of the change in wage. Under these conditions, a negative relation between the quantity of labor supplied and the wage rate is easily understood, but this does not reflect on the slope of the supply curve at a point in time where wealth is constant.

This is not to imply that the aggregate supply curve of labor can never slope downwards or that (in the aggregate) Giffen goods can never be observed. These phenomena are not ruled out on a priori grounds, but may occur only under much more stringent conditions than usually specified. Suppose that the proceeds of a tax imposed on capital are used to subsidize labor. The relative price of labor will increase, and individuals supplying labor will experience positive wealth effects. On this account, the slope of the supply curve may become negative. On the other hand, individuals, in their capacity as owners of capital, will experience negative wealth effects and will demand less leisure and offer more labor services. For an individual whose capital holdings are exactly proportionate to the quantity of labor he supplies, the two income effects will cancel out. The other extreme is where all but one individual have no capital whatsoever while a single individual owns the entire capital stock of the economy. The wealthy owner will increase his labor supply as a result of the decline in his wealth by at most twenty-four hours a day. Since his contribution will be a minute fraction of the total, the aggregate supply curve can easily be negatively sloped. The actual situation is somewhere in between these two extremes and while the notion that the aggregate supply curve may slope down at a point in time is greatly weakened, it cannot be entirely rejected on a priori grounds.

IV. The Empirical Evidence

In *The Price of Leisure*, John Owen compiled *U.S.* time-series spanning the years 1901–61 for (among other things) the number of weekly hours of work and the corresponding real hourly wage. A number of important adjustments to the raw data make these series particularly useful for our purposes, especially in adjusting weekly hours for vacations and holidays and in excluding working students from the series from 1940 on.

The data, however, should be interpreted with great care. The coverage is for "private, non-agricultural wage and salary workers." If we assume, as seems reasonable, that the larger the asset holdings of an individual, the larger the chance that he will become self-employed, then the

[16] By the same token, it is not surprising that Stigler's search for Giffen goods is in vain. See George Stigler.

above classification of workers is biased towards individuals with low asset holdings. This has strong implications in interpreting the time-series. First, we saw that when the wage is parametric, the supply curve of individuals will vary substantially according to whether their level of assets is such that $A-S>0$ or $A-S<0$; this is evident in a comparison of the first and third rows of panels in Figure 1. In particular, as asset holdings increase, the relation $A-S<0$ will change to $A-S>0$, and at the same time, individuals will tend to drop out of the class of workers covered by the series. The individuals covered by the time-series, then, have particular asset holdings that will tend to bias the observed relation towards a negative slope.

Second, as noted above in the section on aggregation, to the extent that the change in wages is not accompanied by a shift in the transformation curve, the income effect on wage earners will be more or less matched by an income effect in the opposite direction on asset owners. But since the latter tend to be excluded from the sample, the effect on them is missed, and again the aggregative curve is biased towards a negative slope.[17]

In addition, during the period covered a dramatic shift occurred in productivity, and the observed increase in wage earners' incomes was not at the expense of asset holders. Consequently, the evidence afforded by the aggregative time-series corresponds more closely to that of the individual where wages are parametric and where increased wages mean a higher level of income than to the aggregative supply curve at a point in time where total income is constant.

The above discussion leads us to view the time-series as being obtained from a population for which asset holdings are not

[17] These two biases, however, are fortunate from the viewpoint of identifying the parameters of the underlying utility function.

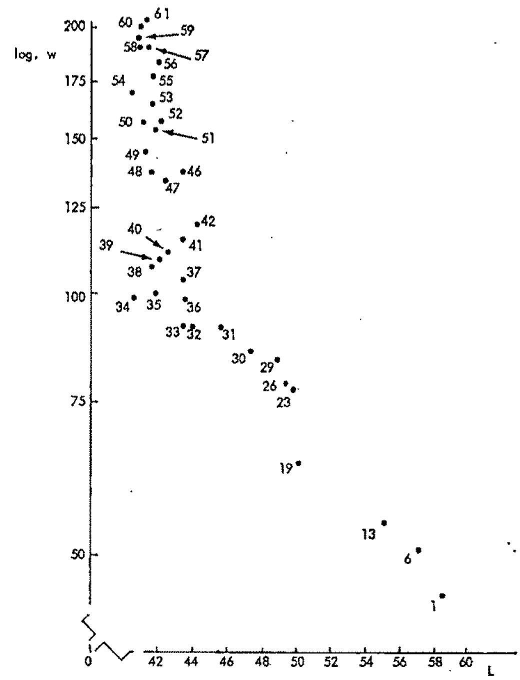


FIGURE 2

high (though perhaps rising) and to view the wage changes as parametric. So the data may narrow down the range of utility functions consistent with the evidence by shedding some light on the constanty of σ and on its value.

The data are plotted in Figure 2, where the horizontal axis measures hours worked per week and the vertical axis measures (on a logarithmic scale) the index of real wages. Individual points are identified according to their year. From 1901 to 1929 only selected points are available; from 1929 to 1961 the data are annual except for the omission of the war years, 1943–45.

The pattern that emerges conforms closely to Panel VIII of Figure 1, where $A-S<0$ and where $\sigma=1$.[18] During the first half of the period, from 1901 to the

[18] We should bear in mind that each of the diagrams in Figure 1 is drawn for a constant level of nonwage income. Since nonwage income is not held constant in the actual data, and since such income may be correlated with the level of wages, a possibility of a bias is present.

mid-1930's, real wages doubled, and weekly hours fell from 58 to about 42. During the second period, from the mid-1930's to 1961, real wages doubled again but hours remained virtually constant, particularly so if the war-related years are ignored. From 1948 to 1961, the range of variation of weekly hours is in the narrow band between 40.4 (1954) and 42 (1952) with no apparent trend.[19] This corresponds to the Cobb-Douglas utility function where the supply curve approaches asymptotically a positive number of hours.[20] Given that American wages in 1901 were substantially above those required for subsistence, the above evidence is insufficient to determine the minimum levels of consumption and rest.

Owen's data go only to 1961 and cover a very broad class of workers. During the 1960's real wages continued to increase, but weekly hours showed very little, if any, decline. Weekly hours for all production workers fell from 38.6 in 1961 to 37.0[21] (the coverage does not differ much from Owen's but the hours are unadjusted). This decline, however, is entirely due to a decline in hours in wholesale and retail trade (from 38.3 to 35.1); none of the other four subclasses, which include mining, construction, manufacturing and finance, showed any decline. Note that both 1961 and 1971 had high unemployment rates. The last decade, then, conforms to the pattern of the earlier period.

An interesting empirical implication (also useful for testing the conclusion that the utility function is Cobb-Douglas) is that as wages go up the divergence with respect to hours of work tends to disappear as between the capitalist and noncapitalist classes, i.e., between those with $A-S>0$ and those with $A-S<0$. Casual notions suggest that working habits of wealthier individuals in the United States do not differ much from those of less well-off individuals. Casual notions also suggest that the "leisure" class is quite prominent in poor countries where the real wages are low. These notions, if correct, tend to support the hypothesis that $\sigma=1$. Of course, a more careful and systematic analysis of the evidence is necessary to determine whether these casual notions are correct.

V. Some Further Comments

The attainable (survival) consumption set used in the preceding analysis was of a simple, "fixed coefficient" nature. A more complicated (and perhaps more realistic) formulation, allowing some substitution between the two commodities in meeting the requisites of survival, would in principle cause few difficulties in the formal theory.[22] However, the notion of net consumption (above survival) would then have little intuitive appeal, and it thus might be impossible to find a formulation of preferences yielding simple expressions like equations (8) and (9) for the slopes of the demand curves. The assumptions used do, however, seem to strike a reasonable balance between complicating the analysis severely and ignoring subsistence needs completely, since they allow simple types of preferences to generate a wide variety of types of demand (supply) curves.

Thus Hanoch verges on an error when he asserts (p. 639) that *CES* utility functions cannot produce backward bending

[19] The range will be further narrowed if we take account of cyclical fluctuations in unemployment, which are negatively correlated with the number of hours.

[20] It is interesting to note that since the end of the Second World War, weekly hours in Great Britain are also almost constant. There, however, the number is around 48 hours per week, The difference of about 7 hours per week between the two countries is an interesting puzzle.

[21] See the *Monthly Labor Review*.

[22] Provided that the new attainable set is convex. Otherwise, it may be impossible to even define convex preferences over the entire set of attainable consumption bundles. See Arrow and Gerard Debreu (1954) and Debreu.

supply curves and that a *C-D* utility function produces a vertical supply curve. His failure to take into account the wealth position and subsistence requirements of the individual has led him to use an unnecessarily complicated utility function to get his desired qualitative results. His illustration of a "conventional" supply curve, with $(A-S)$ implicitly set at zero, shows how unreasonable one can get if minimum consumption is not specified. Since his supply of labor becomes zero at all wage rates below $w=1$,[23] his consumer then would apparently prefer to starve to death rather than work and live at a wage less than 1.

In empirical studies of the supply of labor, the asset position of the individual is seldom taken into account. But a comparison of the first and third rows of Figure 1 clearly shows that even when assets are present we cannot simply add nonwage income as another variable in, say, a regression equation. Clearly, the coefficient for the wage rate depends on the level of nonwage income, and it even changes sign as assets pass through the level $A=S$. A simple relation allowing for no interaction effects, even if statistically "significant," gives average results for diverse groups of individuals, which in this case are entirely meaningless. More specifically, in a sample with assets distributed on both sides of the subsistence level, the estimated effect of the wage rate on the supply of labor will be biased toward zero. Given the model formulated above, however, the correct specification for the labor supply is shown by equations (A 10) and (A 11) of the Appendix. Equation (A 11), where it is assumed that $\sigma=1$, can be readily estimated by standard techniques; (A 10) poses more of a problem but could also be estimated if the hypothesis that $\sigma=1$ is not tenable.

[23] The units of w are not specified; thus, the entry point into the labor market is *not* invariant to the choice of units for the consumption good. His specific utility function in our notation is:

$$V(R, C) = R \exp\left[-\exp\left(-\frac{C}{R}\right)\right]$$

Appendix

To derive equation (10) it will be convenient to solve explicitly for the effects of asset changes on the demands for C and R. We can differentiate equations (4) and (6) totally to get

$$dC + w dR = dA + (D' - R)dw$$

$$f_1 dC + f_2 dR = dw$$

Letting dw be zero and recalling that at equilibrium $w=f$, we see that the partial effect of income on leisure is

$$\frac{\partial R}{\partial A} = f_1/(ff_1 - f_2) \tag{A 1}$$

Substituting this into (9) and rearranging, we have

$$\begin{aligned} \frac{\partial L}{\partial w} &= f_1/(ff_1-f_2)\left[\frac{1}{f_1}-(D'-R)\right] \\ &= \frac{1}{w} f_1/(ff_1 - f_2) \\ &\quad \cdot\left[\frac{f}{f_1}-(C+S-A)\right] \end{aligned} \tag{A 2}$$

where the budget constraint (4) has been used to substitute for $(D'-R)$ and we have again used $w=f$.

Assume now that the individual's preferences are homothetic in C and R, so the marginal rate of substitution is a function only of the ratio of consumption to leisure—that is, $f(C, R)$ is homogeneous of degree zero—so we may write

$$f(C, R) = h\left(\frac{C}{R}\right)$$

Then f_1 and f_2 can be written as

$$f_1 = \frac{h'}{R'}; \quad f_2 = -\frac{C}{R}\frac{h'}{R} \tag{A 3}$$

and we have

$$\text{(A 4)} \quad ff_1 - f_2 = \frac{h'}{R}\,\frac{(C + hR)}{R} > 0$$

by convexity. So it must be true that $h' > 0$. The income effects then become

$$\text{(A 5)} \quad \frac{\partial R}{\partial A} = \frac{R}{(C+wR)} \text{ and } \frac{\partial C}{\partial A} = \frac{C}{(C+wR)}$$

where the derivation of the latter parallels that for leisure.

The elasticity of substitution between the net quantities of consumption and leisure is defined as

$$\sigma = \frac{d\left(\frac{C}{R}\right) \Big/ \left(\frac{C}{R}\right)}{df/f}$$

where the differentials are for movements along an indifference curve. When preferences are homothetic this is simply

$$\text{(A 6)} \qquad \sigma = \frac{Rh}{Ch'}$$

and then we have

$$\text{(A 7)} \qquad \frac{f}{f_1} = \frac{Rh}{h'} = C\sigma$$

Equations (A5) and (A7) may then be substituted into (A2) to yield (A8).

$$\text{(A 8)} \quad \frac{\partial L}{\partial w} = \frac{1}{w}\,\frac{R}{(C + wR)} \cdot [C(\sigma - 1) + (A - S)]$$

Working in a parallel manner, equation (7) can be rewritten as

$$\text{(A 9)} \quad \frac{\partial C}{\partial w} = \frac{C}{(C + wR)}\,[R(\sigma - 1) + D']$$

It can also be shown that no weaker specifications on the individual's preferences will give these results.

The further assumption that the elasticity of substitution is a constant yields the explicit supply function

$$\text{(A 10)} \quad L = \frac{\left(\frac{\alpha}{1-\alpha}\right)^{\sigma} w^{\sigma} D' - (A - S)}{\left(\frac{\alpha}{1-\alpha}\right)^{\sigma} w^{\sigma} + w}$$

which reduces to

$$\text{(A 11)} \quad L = \alpha D' - (1 - \alpha)\left(\frac{A - S}{w}\right)$$

when $\sigma = 1$. This is the supply function that would be generated by a Cobb-Douglas type of utility function.

REFERENCES

K. J. Arrow and G. Debreu, "Existence of an Equilibrium for a Competitive Economy," *Econometrica*, July 1954, *22*, 265–90.

M. J. Bailey, "The Marshallian Demand Curve," *J. Polit. Econ.*, June 1954, *62*, 255–61.

———, *National Income and the Price Level*, New York 1962, p. 34.

K. E. Boulding, *Economic Analysis*, 3d ed., New York 1955.

G. Debreu, *Theory of Values: An Axiomatic Analysis of Economic Equilibrium*, New York 1959.

M. Friedman, *Price Theory*, rev., Chicago 1966, p. 204.

———, "The Marshallian Demand Curve," *J. Polit. Econ.*, Dec. 1949, *57*, 463–94; reprinted in *Essays in Positive Economics*, Chicago 1953.

G. Hanoch, "The 'Backward-Bending' Supply of Labor," *J. Polit. Econ.*, Dec. 1965, *73*, 636–42.

J. R. Hicks, *Value and Capital*, 2d ed., Oxford 1946, p. 37.

L. R. Klein and H. Rubin, "A Constant Utility Index of the Cost of Living," *Rev. Econ. Stud.*, 1948–49, *15*, 84–87.

J. D. Owen, *The Price of Leisure*, Rotterdam 1969.

L. Robbins, "On the Elasticity of Demand for Income in Terms of Effort," *Economica*, June 1930, *29*, 123–29.

P. A. Samuelson, "Some Implications of

'Linearity'," *Rev. Econ. Stud.*, 1948–49, *15*, 88–90.
———, *Economics*, 8th ed., New York 1970.
G. J. Stigler, "Notes on the History of the Giffen Paradox," *J. Polit. Econ.*, Apr. 1947, *55*, 152–56.
R. D. Stone, "Linear Expenditure Systems and Demand Analysis: An Application to the Pattern of British Demand," *Econ. J.*, Sept. 1954, *64*, 511–27.
Monthly Labor Review, U.S. Department of Labor, July 1972, p. 82.

The Intrafamily Allocation of Time: The Value of the Housewives' Time

By REUBEN GRONAU*

In the new approach to consumption theory, the consumption activity is regarded as a production process in which time and goods are combined to produce utility. This new theory has revived interest in the family as the basic consumption unit.[1] While classical theory regarded "household" as synonymous with "individual," the new approach came to recognize that the members of a family each play a different role in the production of utility. Various authors[2] suggested that the classical dichotomy of "work in the market" versus "leisure" may serve as a good approximation of the role the husband plays in the production activity of the household but does gross injustice to the wife. To call the whole of the time spent by the wife outside the market sector "leisure" is to overlook the production activities she engages in at home. These activities are better termed work at home, and the wife's allocation of time should therefore be analyzed in terms of a three-way division of work in the market, work at home and leisure.

Recently, there have been some attempts to formalize the family decision-making process and to analyze the factors determining the intrafamily allocation of time and goods (Stuart Altman and Robert Barro, Reuben Gronau (1970b), Haim Ofek, James Smith). However, most of these attempts were based on a specific formulation of the form of the family utility function,[3] so that the validity of their results depends on the mathematical form used. A more serious limitation of most of these models (Altman and Barro, Ofek) is the failure to relate to or explain the salient feature of the intrafamily allocation of time—the fact that at each point of time over 60 percent of all married women are not actively engaged in market production and that during any given year not more than 50 percent participate in the labor force. The explicit or implicit assumption that the wife works in the market limits the relevance of these models to at most one-half of all families, but leaves unanswered questions relating to the remainder, where the wife's sole occupation is housewife.

This paper attempts a general formulation of the intrafamily allocation of time. The members of a family allocate their time according to their comparative advantage in the production of market and

* Hebrew University, Jerusalem. Parts of this paper are based on my study, "The Labor Force Participation of Israeli Women," carried out at the Maurice Falk Institute for Economic Research. I finished the paper while a postdoctoral fellow at the University of Chicago. I acknowledge the financial support of the Rockefeller Foundation through a grant to the University of Chicago for research on the Economics of Population and Family Decision Making. I benefitted from the suggestions of Gary Becker, Giora Hanoch, Ruth Klinov, David Levhari, Gregg Lewis, and Shlomo Yizhaki, and I am grateful to Randall Olson for his devoted research assistance.

[1] This new approach is associated with the work of Gary Becker (1965) and Kelvin Lancaster (1966) but can actually be dated back to Wesley Mitchell (1937). Among its recent exponents are Stuart Altman and Robert Barro (1970), Gilbert Ghez and Becker (1972), Gronau (1970), Michael Grossman (1972), Robert Michael (1973), Haim Ofek (1970), and others.

[2] Most notably Jacob Mincer (1962, 1963), and later Glen Cain (1966) and Marvin Kosters (1963).

[3] Ofek and Gronau (1970b) used a two-stage *CES* function while Altman and Barro used a Cobb-Douglas function.

home goods. Comparative advantage is in turn determined by their relative wage rates and their efficiency in the production of home goods. Husbands' wage rates as a rule exceed those of wives. Thus, we observe that, in general, husbands specialize in work in the market, while the wife specializes in the production of home goods. Very crudely, one can distinguish three situations of specialization: the case where the husband divides his time between work in the market and leisure and the wife works both in the market and at home, spending the rest of her time on leisure; the case where the wife drops out of the labor force; and the case where the husband enters the home production process. These three phases are reflected in a difference in the factors determining the value of the wife's time. The value placed by the family on the wife's time while she works in the market is, of course, determined by her own marginal wage rate. When she leaves the labor market this tie is severed, and the price of her time is determined, in the second case, by family income, and, in the third case, it is determined by her husband's wage rate.

The wife's decision whether to participate in the labor force can, therefore, be viewed in terms of the comparison between her value of time in the absence of market opportunities and her potential wage rate. The result of this comparison is recorded in the labor force statistics. These statistics yield themselves however to two different interpretations, based on the self-selection of working wives from a frequency distribution of their alternative price of time. According to the first, the wives who work are those who are the least productive in the home sector, i.e., those whose value of time is the lowest, and, hence, the mean price of time of housewives exceeds the average wage rate of working women. By the second interpretation, those who work are the women most fit for market work, i.e., those who have received the highest wage offers, and, hence, the mean price of time falls short of the average wage rate.

Using 1960 *U.S.* data it is found that under the first assumption, the housewives' average price of time exceeds the average wage of working women by less than 20 percent. Given the second assumption, the housewives' average price of time falls short of the average wage rate by about 20–30 percent. White women assign to their time a higher value than nonwhites. Only part of this difference may be explained by income differentials, the rest arising, seemingly, from differences in their age-education composition. Finally, it is found in the case of nonwhite women that the existence of a young child (less than three years old) raises his mother's price of time by about 6–8 percent. We could not, however, derive a similar estimate for white women.

I. The Intrafamily Allocation of Time

Let us consider for simplicity a household (family) consisting of two members: husband and wife. The household combines its members' leisure time with market and home goods to generate utility (U).

$$U = U(M, H, L_1, L_2) \tag{1}$$

The arguments of the utility function will be termed factors, where M denotes the amount of market goods, H the amount of home goods, L_1 the amount of the husband's leisure, L_2 the amount of leisure enjoyed by the wife.[4] Home goods, in turn, are produced with a combination of market inputs and time and can be produced by either husband or wife.

One of the things distinguishing a family

[4] This formulation ignores the effects that intrafamily distribution of goods may have on family welfare and assumes implicitly that work, whether in the market or in the nonmarket sector, carries no utility (or disutility).

from a single-person household is the exchange taking place within the family. Thus, the husband does not have to rely on his own talents to produce these goods (say, a meal) but can muster the services of his wife in exchange for say, his working in the market. Thus, $H = H_1 + H_2$ where

$$(2) \quad H_i = H_i(X_i, T_{Hi}), \qquad i = 1, 2$$

H_i being the amount of home goods produced by person i, and X_i and T_{Hi} being, respectively, the amount of market inputs and time used in the process. The family aims at maximizing its utility subject to the constraints it confronts. These constraints are of two kinds: (a) the budget constraint, and (b) the time constraint.

The family pools its pecuniary resources. Adopting a one-period model, the budget constraint states that expenditure on market goods and inputs cannot exceed the family income (I).

$$(3) \quad P_M M + P_X X = W_1 T_{M_1} + W_2 T_{M_2} + V = I$$

where W_i denotes the wage rate of member i, T_{Mi} the amount of time he spends working in the market, V other sources of income besides earnings, P_M and P_X the price of market goods (M) and inputs (X), respectively, and $X = X_1 + X_2$.

The family faces two separate time constraints stating that the amount of time spent by person i on work in the market, work at home, and leisure cannot exceed the total amount of time available (T_0)

$$(4) \quad T_{Mi} + T_{Hi} + L_i = T_0 \qquad i = 1, 2$$

The maximization of the utility function (1) subject to the production function for home goods (2), and the budget and time constraints ((3) and (4), respectively), yields the family members' optimum allocation of time and the family's optimum allocation of expenditure between market goods and market inputs.

This optimum solution depends on the specific nature of the utility function, the home goods production functions and the family members' wage rates and other sources of income. Let us assume, for simplicity, that the production of home goods calls for the combination of time and market inputs in fixed proportions.[5] Moreover, let us assume that these functions differ between husband and wife:

$$(5) \quad H_i = \min\left(\frac{X_i}{\gamma_i}, \frac{T_{Hi}}{\delta_i}\right) \qquad i = 1, 2$$

$1/\gamma_i$ and $1/\delta_i$ being, respectively, the marginal products of market inputs and time in the production of home goods. Given this assumption, inputs vary proportionately with the amount of home goods produced

$$(6) \quad \begin{aligned} X_i &= \gamma_i H_i \\ T_{Hi} &= \delta_i H_i \qquad i = 1, 2 \end{aligned}$$

The marginal cost of producing one unit of home goods by person $i(\Pi_i)$ is

$$(7) \quad \Pi_i = \gamma_i P_X + \delta_i W_i^* \qquad i = 1, 2$$

where W_i^* is the value placed on the time of person i.

If we assume that the values placed on the husband's and wife's time are constant (though not identical) and do not vary with amount of home goods produced,[6] and given our previous assumption concerning the constancy of δ_i and γ_i the marginal costs of producing home goods are constant and differ, in general, between husband and wife. The family will turn for its supply of home goods to the cheaper of the two producers.

[5] This strong assumption could be replaced by the weaker assumption that the production function of home goods is homogeneous of degree $t \geq 1$ without affecting the major conclusions of the model but at the cost of greatly complicating the computations.

[6] This assumption is removed later in this section.

Case I: Both Members of the Family Participate in the Labor Force

If both members of the family participate in the labor force, the value of their time equals their wage rate ($W_i^* = W_i$).[7] Their marginal costs of producing home goods are therefore

$$\Pi_i = \gamma_i P_X + \delta_i W_i \qquad i = 1, 2 \tag{8}$$

The ith person's marginal costs of producing home goods depend on his marginal productivity (γ_i and δ_i) and his wage rate. Since in general the husband's exceeds the wife's wage rate the price charged by the husband for home goods exceeds that charged by the wife unless this difference in wages is offset by differences in efficiency. Consequently, all home goods are going to be produced by the wife, while her husband divides his time between work in the market and leisure (i.e., $T_{H_1} = 0$).

In this case, one can incorporate both time constraints (4) into the budget constraint (3) to obtain one ultimate constraint

$$\begin{aligned} I &= P_M M + P_X X_2 \\ &= W_1(T_0 - L_1) + W_2(T_0 - T_{H_2} - L_2) + V \end{aligned} \tag{9}$$

Inserting the values of X_2 and T_{H_2} in terms of the output of home goods into the equation, this constraint can be written

$$\begin{aligned} P_M M + \Pi H + W_1 L_1 + W_2 L_2 \\ = (W_1 + W_2) T_0 + V = I^* \end{aligned} \tag{10}$$

where $\Pi = \gamma_2 P_X + \delta_2 W_2$ is the price of home goods, and I^* is the family's "full income," i.e., the income the family could have earned had it devoted all its time to work in the market.

The maximization of utility subject to the ultimate constraint (10) yields the familiar optimum conditions

$$\begin{aligned} u_M &= \lambda P_M \\ u_H &= \lambda \Pi \\ u_{L_1} &= \lambda W_1 \end{aligned} \tag{11}$$

and

$$u_{L_2} = \lambda W_2$$

where $u_Z(Z{:}M, H, L_1, L_2)$ denotes the marginal utility of factor Z, and λ is the marginal utility of income.

Given these necessary conditions one can generate the demand elasticities for market goods, home goods, and leisure. Of particular interest are the demand elasticities with respect to the husband's and wife's wage rates. The demand elasticity for factor $Z(Z{:}M, H, L_1$ or $L_2)$ with respect to the husband's wage rate (ϵ_{ZW_1}) is

$$\epsilon_{ZW_1} = \frac{W_1 L_1}{I^*} \sigma_{L_1 Z} + \frac{W_1 T_{M_1}}{I^*} \epsilon_{ZI^*} \tag{12}$$

where $\sigma_{L_1 Z}$ denotes the Allen partial elasticity of substitution between factor Z and the husband's leisure, and ϵ_{ZI^*} denotes the "full income" elasticity of demand.[8] Specifically, the demand elasticity for the husband's leisure with respect to his wage rate is shown by equation (13).

[7] Recall our assumption that work does not carry any utility (see fn. 4). Without this assumption one has to adjust the value of time for the money equivalent of the marginal utility (or disutility) of work (see Bruce Johnson (1966) and Gronau (1970a)).

[8] The full income elasticity exceeds the income elasticity of demand

$$\epsilon_{ZI^*} = \frac{I^*}{I} \epsilon_{ZI} > \epsilon_{ZI},$$

and equation (12) can also be written

$$\epsilon_{ZW_1} = \frac{W_1 L_1}{I^*} \sigma_{L_1 Z} + \frac{W_1 T_{M_1}}{I} \epsilon_{ZI}$$

When the utility function is linear homogeneous $\epsilon_{ZI^*} = 1$ and

$$\epsilon_{ZW_1} = \frac{W_1 L_1}{I^*} \sigma_{L_1 Z} + \frac{W_1 T_{M_1}}{I^*}$$

$$(13)\qquad \epsilon_{L_1W_1} = \frac{W_1L_1}{I^*}\sigma_{L_1W_1} + \frac{W_1T_{M_1}}{I^*}\epsilon_{L_1I^*}$$

The husband's elasticity of supply of working time is therefore

$$(14)\qquad \epsilon_{T_{M_2}W_1} = -\frac{L_1}{T_{M_1}}\epsilon_{L_1W_1}$$

$$= -\left(\frac{W_1L_1}{I^*}\frac{L_1}{T_{M_1}}\sigma_{L_1L_1} + \epsilon_{L_1I^*}\right)$$

Since $\sigma_{L_1L_1}$ is always negative and $\epsilon_{L_1I^*}$ can be safely assumed to be positive, we encounter the familiar result that the slope of the labor supply curve depends on the relative magnitudes of the elasticity of substitution and the full income elasticity of leisure.[9]

Similarly, the cross elasticities of demand for home goods and the wife's leisure with respect to the husband's wage rate are

$$(15)\qquad \epsilon_{HW_1} = \frac{W_1L_1}{I^*}\sigma_{L_1H} + \frac{W_1T_M}{I^*}\epsilon_{HI^*}$$

$$\epsilon_{L_2W_1} = \frac{W_1L_1}{I^*}\sigma_{L_1L_2} + \frac{W_1T_{M_1}}{I^*}\epsilon_{L_2I^*}$$

The elasticity of demand for the wife's working time at home with respect to her husband's wage rate ($\epsilon_{T_{H_2}W_1}$) equals the demand elasticity for home goods with respect to $W_1(\epsilon_{HW_1})$. If home goods and the wife's leisure are competitive with the husband's leisure (i.e., if $\sigma_{L_1L_2}$, $\sigma_{L_1H}>0$), an increase in the husband's wage rate will be accompanied by an increase in the amount of time the wife spends at home and, consequently, by a decline in the amount of time she devotes to the market sector.[10] This is seen in equation (16).

$$(16)\qquad \epsilon_{T_{M_2}W_1} = -\left\{\frac{W_1L_1}{I^*}\left(\frac{L_2}{T_{M_2}}\sigma_{L_1L_2} + \frac{T_{H_2}}{T_{M_2}}\sigma_{L_1H}\right) + \frac{W_1T_{M_1}}{I^*}\left(\frac{L_2}{T_{M_2}}\epsilon_{L_2I^*} + \frac{T_{H_2}}{T_{M_2}}\epsilon_{HI^*}\right)\right\}$$

An increase in the wife's wage rate increases both the price of her leisure and the price of home goods, the increase in the latter being a function of the share of the cost of time in the total production cost of home goods. The demand elasticity of factor $Z(Z\!:\!M$, H, L_1, or $L_2)$ with respect to the wife's wage rate reflects the substitution elasticity of factor Z for both the wife's leisure and home goods, as well as the full income elasticity of this factor

$$(17)\qquad \epsilon_{ZW_2} = \frac{W_2L_2}{I^*}\sigma_{L_2Z} + \frac{W_2T_{H_2}}{I^*}\sigma_{HZ} + \frac{W_2T_{M_2}}{I^*}\epsilon_{ZI^*}$$

Under the previous assumption that the husband's leisure is competitive with both home goods and the wife's leisure one would expect the husband's leisure to increase, and hence his supply of labor to decrease, as his wife's wage rate increases. This is seen in equation (18).

[9] When the utility function is linear homogeneous the necessary condition for the backward bending supply curve of labor is

$$|\sigma_{L_1L_1}| > \frac{T_{M_1}}{L_1} = \frac{T_{M_1}}{T_0 - T_{M_1}}$$

[10] This conclusion holds even if the husband's and wife's leisure are complementary factors and $\sigma_{L_1L_2}<0$ so long as the elasticity of substitution between the husband's leisure and home goods (σ_{L_1H}) is positive and sufficiently large.

$$(18)\quad \epsilon_{TM_1W_2} = -\frac{L_1}{T_{M_1}}\left(\frac{W_2L_2}{I^*}\sigma_{L_1L_2} + \frac{W_2T_{H_2}}{I^*}\sigma_{L_1H} + \frac{W_2T_{M_2}}{I^*}\epsilon_{L_1I^*}\right)$$

However, it can be shown that, in general, the wife's supply of labor is more sensitive to changes in her husband's wage rate than her husband's supply of labor is to changes in her wage rate.[11]

The elasticity of the wife's supply of labor with respect to her own wage rate depends on the family's demand for her leisure and for home goods. These elasticities are

$$(19)\quad \epsilon_{L_2W_2} = \frac{W_2L_2}{I^*}\sigma_{L_2L_2} + \frac{W_2T_{H_2}}{I^*}\sigma_{HL_2} + \frac{W_2T_{M_2}}{I^*}\epsilon_{L_2I^*}$$

and

$$\epsilon_{HW_2} = \frac{W_2L_2}{I^*}\sigma_{HL_2} + \frac{W_2T_{H_2}}{I^*}\sigma_{HH} + \frac{W_2T_{M_2}}{I^*}\epsilon_{HI^*}$$

The sign of these elasticities is indeterminate since the income and substitution effects tend to have opposite effects on the quantity of leisure demanded.[12] Given the income elasticities ($\epsilon_{L_2I^*}$ and ϵ_{HI^*}) and the substitution elasticities ($\sigma_{L_2L_2}$, σ_{HH}, and σ_{HL_2}), the demand elasticities ($\epsilon_{L_2W_2}$ and ϵ_{HW_2}) depend on the distribution of the wife's time between market and nonmarket activities. The smaller amount of time the wife spends in the market, the smaller the income effect, and since

$$(20)\quad \epsilon_{T_{M_2}W_2} = -\left(\frac{L_2}{T_{M_2}}\epsilon_{L_2W_2} + \frac{T_{H2}}{T_{M_2}}\epsilon_{HW_2}\right)$$

the greater the tendency of her supply of labor to be positively sloped. Furthermore, if the income elasticities of all factors are the same (specifically, if $\epsilon_{L_1I^*} = \epsilon_{L_2I^*} = \epsilon_{HI^*}$) one would, from a comparison of (20), (19), and (14), expect the wife's labor supply curve to have less tendency to bend backwards than her husband's, since the wife's earnings constitute a smaller share of full income than the husband's. Put differently, if the wife's labor supply curve bends backwards it will do so only at a point (i.e., number of hours) which is to the right of the bending point of her husband's supply curve.

Case II: Only the Husband Participates in the Labor Market

An increase in family income, whether as a result of an increase in the husband's wage rate (W_1) or in other sources of income (V), is accompanied by an increase

[11] Comparing (16) and (18) we get

$$\epsilon_{T_{M_2}W_1} - \epsilon_{T_{M_1}W_2} = \frac{1}{I^*}\left\{\frac{L_1(W_1T_{M_1} - W_2T_{M_2})}{T_{M_1}T_{M_2}}(L_2\sigma_{L_1L_2} + T_{H_2}\sigma_{L_1H}) + \left(\frac{W_1T_{M_1}}{T_{M_2}}(L_2\epsilon_{L_2I^*} + T_{H_2}\epsilon_{HI^*}) - \left(\frac{W_2T_{M_2}}{T_{M_1}}L_1\epsilon_{L_1I^*}\right)\right\}$$

If the full income elasticities $\epsilon_{L_1I^*}, \epsilon_{L_2I^*}$ and ϵ_{HI^*} are about equal, $\epsilon_{T_{M2}W_1} > \epsilon_{T_{M1}W_2}$, since the husband spends more time than his wife in the market and earns more than she does.

[12] Assuming home goods and the wife's leisure to be normal (i.e., not inferior) inputs and σ_{ML_1}, σ_{MH}, $\sigma_{L_1L_2}$, $\sigma_{L_1H} > 0$.

in the demand for both wife's leisure and home goods. Eventually, the wife drops out of the labor market altogether, dividing her time exclusively between work at home and leisure. The wife's exit from the labor market severs the link between the value of her time and her potential wage rate (i.e., the wage she could have earned had she stayed in the labor force). The value placed by the family on the wife's time depends on the value of the marginal product of her labor inputs in the production of home goods. The mere fact that the wife prefers not to work in the market indicates that the value of her time exceeds her potential wage rate.

Formally, when the wife is not working in the market one cannot treat identically the two separate time constraints confronting the family, i.e., one cannot incorporate both constraints in the ultimate budget constraint. The family's budget is unaffected by the wife's allocation of time between work at home and leisure. If we still maintain that the husband does not work at home, one can write the budget constraint as

$$P_M M + P_X X_2 = W_1 T_{M_1} + V = W_1(T_0 - L_1) + V \tag{21}$$

or alternatively as

$$P_M M + P_X \gamma_2 H + W_1 L_1 = W_1 T_0 + V \tag{22}$$

The family, however, faces an additional constraint—the wife's time constraint, stating that the time used by the wife in the production of home goods plus her leisure time cannot exceed total time available.

$$T_{H_2} + L_2 = \delta_2 H + L_2 = T_0 \tag{23}$$

The maximization of utility therefore takes place under two separate constraints: the budget constraint (22) and the wife's time constraint (23).

The optimum conditions look familiar

$$\begin{aligned} u_M &= \lambda P_M \\ u_H &= \lambda \Pi^* \\ u_{L_1} &= \lambda W_1 \end{aligned} \tag{24}$$

and

$$u_{L_2} = \lambda W_2^*$$

where $\Pi^* = \lambda_2 P_X + \delta_2 W_2^*$. However, in this case, the value W_2^* placed on the wife's time, and, consequently, the price of home goods Π^* is not exogeneously given, but rather determined by the maximization process. A change in the parameters (P_M, P_X, W_1, and V) changes both the optimum solution and the price placed by the family on the wife's time and home goods.

For example, an increase in the family's other sources of income (V) would, in the absence of the wife's time constraint, result in an increase of all the four components of the utility function (assuming none of the four is an inferior input). However, given the time constraint the wife cannot simultaneously increase both her leisure and the time she spends in producing home goods. The increase in the demand for the wife's time results in an increase in the price placed on this time. The increase in income is, therefore, accompanied by a substitution of the husband's leisure and market goods for the wife's leisure and home goods.[13] The leisure of husbands whose wives do not work should, therefore, be more sensitive to changes in income than the leisure of husbands whose wives work in the market. Put differently, the tendency for the husband's supply of labor to bend backwards should be more prevalent among husbands

[13] The proof is included in an unpublished Appendix. The result that the price assigned to the housewife's time increases with family income is a general one, and holds even if one does not distinguish between the wife's leisure and her work at home. The distinction is, however, crucial if one wants to explain the different participation patterns of husbands and wives.

of nonworking wives than among husbands of working wives.

As for the wife, an increase in income results in an increase in the value of her time and an increase in the price of both her leisure and home goods, the increase in the latter being a function of the share of the cost of time in the total production costs of home goods. The demand for the wife's leisure and time inputs in the production of home goods is, therefore, affected by two contradictory forces: the income effect and the substitution effect. Since the total time available for the two activities remains constant, the increase in the price of time must be sufficiently large for the substitution effect of one of these factors to outweigh its income effect. Which of these factors, the wife's leisure or her working time at home, will decline depends on the part time plays in the production of home goods, the income elasticities of the two factors and the substitution elasticities between these factors and market goods and the husband's leisure σ_{MH}, σ_{ML_2}, σ_{HL_1}, and $\sigma_{L_2L_1}$.[14]

Finally, if the husband's wage rate W_1 and other sources of income are sufficiently large, the value placed on the wife's time becomes so large as to make husband-produced and wife-produced home goods equally costly

$$(25)\quad \Pi_1=\gamma_1 P_X+\delta_1 W_1=\gamma_2 P_X+\delta_2 W_2^*=\Pi_2^*$$

In this case the husband is called upon to help his wife at home. The husband divides his time between work in the market, work at home, and leisure, while his wife divides her time between work at home and leisure. The wife's value of time is pegged to her husband's wage rate.

$$(26)\quad W_2^*=\left(\frac{\gamma_1-\gamma_2}{\delta_2}\right)P_X+\left(\frac{\delta_1}{\delta_2}\right)W_1$$

[14] The proof is included in an unpublished Appendix.

When husband and wife are equally efficient in the use of market inputs ($\gamma_1=\gamma_2$) the wife's value of time is proportionate to the husband's. The wife's price of time will equal the husband's wage rate if both husband and wife share the same production function.

In this case an increase in other sources of income (V), husband's wage rate remaining constant, does not change the relative prices of goods and leisure. Thus, the increase in V should result in an increase in the demand of all factors. Both husband's and wife's leisure are expected to increase. Likewise, one expects the production of home goods to expand, the husband taking over some of the work given up by his wife. The increase in the husband's leisure and his work at home result in a decline of the amount of work supplied in the market.

An increase in the husband's wage rate raises the price of both his and his wife's leisure and the price of home goods relative to the price of market goods. The substitution effect tends to increase the husband's work in the market while the income effect works in the opposite direction. Similarly, one cannot predict the effect of an increase in the husband's wage rate on the wife's allocation of time. It depends on the relative change of the price of the wife's leisure versus the price of home goods and on the income and substitution elasticities of these two factors.

II. The Value of the Housewives' Time

An empirical estimation of the demand for leisure and the supply of work at home calls for detailed data concerning the time budgets of the various family members. The existing published data are too crude to provide conclusive results. In the absence of data on the nonmarket sector one has to make inferences about the decision-making process of the family from its revealed preferences with respect to work in

the market. This subject has been investigated in the family context, with somewhat mixed results.[15] However, one implication of our model that has been entirely overlooked is the implication of labor force participation on the value of housewives' time.

Over 60 percent of married women in the United States are classified as full-time housewives (William Bowen and T. A. Finegan, p. 88). Thus, the wage rate serves only as a very crude indication of the value of time for more than one-quarter of the adult population. Given the importance of the price of time as a determinant of housewives' purchasing, travelling, and recreation habits it is of the utmost importance to get some better estimate of the value of time for this population group. Such an estimate can be generated from the observed data on labor force participation.

One of the implications of the preceding section states that while the value of time of working women equals their wage rate, the value placed on the time of housewives exceeds their potential wage rate and increases with family income. The wife's decision to enter the labor market can therefore serve as an indication that in the absence of market opportunities her value of time would have fallen short of her current wage rate. Similarly, the wife's decision to refrain from entering the labor force indicates that her value of time exceeds her potential wage rate.

The rate of labor force participation depends on the joint distribution of the potential wage rate and the wives' price of time in the absence of market opportunities. Formally, let $f(W, W^*)$ be the joint density function of the potential wage W and the price of time W^*, then the participation rate equals

$$(27)\quad P = \text{Prob}\,(W > W^*) = \int_{-\infty}^{\infty}\int_{W^*}^{\infty} f(W, W^*)dWdW^*$$

The average wage of working women and the housewives' average price of time are means of truncated distributions. The average wage $\overline{W}$ equals the expected value of W, where W exceeds W^*,

$$(28)\quad \overline{W} = E(W \mid W > W^*) = \frac{1}{P}\int_{-\infty}^{\infty}\int_{W^*}^{\infty} Wf(W, W^*)dWdW^*$$

and the housewives' average price of time $\overline{W}^*$ is the conditional expectation of W^* where W^* exceeds W.

$$(29)\quad \overline{W}^* = E(W^* \mid W^* > W) = \frac{1}{1-P}\int_{-\infty}^{\infty}\int_{W}^{\infty} W^*f(W, W^*)dW^*dW$$

A prerequisite for the estimation of $\overline{W}^*$ (and the factors determining labor force participation) is the knowledge of the shape and parameters of $f(W, W^*)$.[16]

To estimate $f(W, W^*)$ one has to rely on knowledge of the truncated wage distribution $g(W \mid W > W^*)$ and the portion this distribution constitutes of the total wage offer distribution, i.e., the participation rate $P = \text{Prob}\,(W > W^*)$. In general one does not possess any information with regard to $g(W \mid W > W^*)$ beyond the knowledge of its mean. Thus, one has to replace the missing information by a set of assumptions. I shall assume that W and W^* are independently distributed and that their joint distribution is bivariate normal

[15] Kosters estimated the supply of hours of work of men aged 50–64, but was unable to produce a positive compensated wage rate elasticity. Cain came up with estimates of elasticities which have the right sign. However, a recent paper by Yoram Ben-Porath casts doubts on the validity of Cain's interpretation.

[16] For a more detailed analysis of the problems involved in the estimation of the determinants of labor force participation see Ben-Porath and, in particular, Gregg Lewis.

$$(30)\quad f(W, W^*) = \frac{1}{2\Pi\sigma_W\sigma_{W^*}} \exp\left\{-\frac{1}{2}\left(\frac{W-\mu_W}{\sigma_W}\right)^2 + \left(\frac{W^*-\mu_{W^*}}{\sigma_{W^*}}\right)^2\right\} = \frac{1}{2\Pi\sigma_W\sigma_{W^*}} \cdot \exp\left\{-\frac{1}{2}(x^2+y^2)\right\}$$

where μ_W and μ_{W^*} are the mean values, and σ_W and σ_{W^*} are the standard deviations of the marginal distributions of W and W^*, respectively, and where $x=(W-\mu_W)/\sigma_W$ and $y=(W^*-\mu_{W^*})/\sigma_{W^*}$ are standardized normal variables. Of the two assumptions, normality and independence, the latter is clearly the more controversial since it asserts that none of the factors determining W (for example, education, age, natural ability) affects W^*, and vice versa. This assumption is, therefore, adopted with mixed feelings, being crucial to the estimation procedure.

Let it be assumed that the mean wage rate μ_W is a sole function of age and education and that the mean price of time μ_{W^*} depends solely on the family's income. The rate of labor force participation within a given age-education-income group equals

$$(31)\quad P = \text{Prob}\,(W=\mu_W+x\sigma_W>\mu_{W^*}+y\sigma_{W^*} = W^*) = \text{Prob}\,(x>A+By=y^*) = \frac{1}{2\Pi}\int_{-\infty}^{\infty}\int_{y^*}^{\infty}\exp\left\{-\tfrac{1}{2}(x^2+y^2)\right\}dxdy$$

where $A=(\mu_{W^*}-\mu_W)/\sigma_W$ and $B=\sigma_{W^*}/\sigma_W$. The average wage rate of working women equals

$$(32)\quad \overline{W} = \mu_W + \bar{x}\sigma_W$$

where[17]

[17] See the mathematical appendix of Gronau (1973).

$$(33)\quad \bar{x} = E(x \mid x > y^*) = \frac{1}{2\Pi P}\int_{-\infty}^{\infty}\int_{y^*}^{\infty} x \exp\left\{-\frac{1}{2}(x^2+y^2)\right\}dxdy = \frac{1}{P}\left(2\Pi\frac{\sigma_W^2+\sigma_{W^*}^2}{\sigma_W}\right)^{-1/2} \cdot \exp\left\{-\frac{1}{2}\frac{(\mu_{W^*}-\mu_W)^2}{\sigma_W^2+\sigma_{W^*}^2}\right\}$$

The knowledge of P and $\overline{W}$ is insufficient for the estimation of the four parameters of $f(W, W^*)$, namely μ_W, μ_{W^*}, σ_W and σ_{W^*}. To estimate these parameters we have to adopt two alternative extreme assumptions:

a) $\sigma_W=0$ and σ_{W^*} is insensitive to income

and b) $\sigma_{W^*}=0$ and σ_W is independent of age and education.

If it is assumed that all women in a given age-education group anticipate the same wage rate μ_W (i.e., $\sigma_W=0$), differences in participation behavior of women sharing the same market characteristics are explained in terms of differences in their price of time (see Figure 1). The labor force participation rate within an age-education-income group equals

$$(34)\quad P = \text{Prob}\,(W^*<\mu_W) = \text{Prob}\left(y=\frac{W^*-\mu_{W^*}}{\sigma_{W^*}}<\frac{\mu_W-\mu_{W^*}}{\sigma_{W^*}} = -\frac{A}{B}=Z\right)$$

Moreover, since $\sigma_W=0$, the average wage rate of working women equals the mean value of the wage offer distribution $\overline{W}=\mu_W$. Thus

$$(35)\quad Z = -\frac{A}{B} = \frac{\mu_W-\mu_{W^*}}{\sigma_{W^*}} = \frac{\overline{W}-\mu_{W^*}}{\sigma_{W^*}}$$

or alternatively

$$(36)\quad \overline{W} = \mu_{W^*} + Z\sigma_{W^*}$$

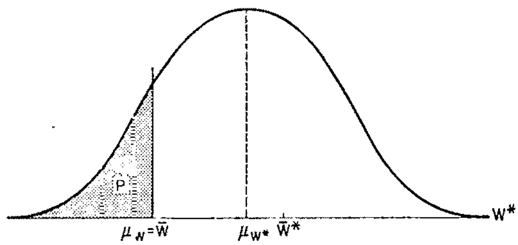


FIGURE 1

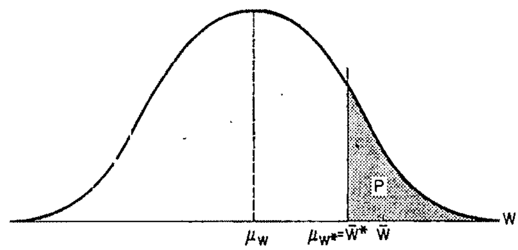


FIGURE 2

Observing that in income group i, P_{ij} percent of the women belonging to potential wage group j (i.e., age-education group j) participate in the labor force, one can (using the tables of the normal distribution) generate the values of Z_{ij} satisfying Prob $(Z<Z_{ij})=P_{ij}$. Given a sufficient number of potential wage groups one can estimate within each income group i

$$(37) \qquad \overline{W}_{ij} = a_i + b_i Z_{ij}$$

the constant term a_i serving as the estimate of the mean value of time μ_{W^*i} in this income group and the regression coefficient b_i serving as an estimate of the standard deviation σ_{W^*i}.

Alternatively, one can assume that differences in participation behavior originate in differences in wage offers, i.e., the standard deviation of the value of time distribution within a given income group equals zero ($\sigma_{W^*}=0$, see Figure 2). The rate of participation within a given age-education-income group is

$$(38) \qquad P = \text{Prob}\,(W > \mu_{W^*}) = \text{Prob}\left(x = \frac{W-\mu_W}{\sigma_W} > \frac{\mu_{W^*}-\mu_W}{\sigma_W} = A\right)$$

By equations (32) and (33) the average wage of working women is $\overline{W}=\mu_W+\bar{x}\sigma_W$ where

$$(39) \qquad \bar{x} = E(x \mid x > y^* = A) = \frac{1}{P\sqrt{2\Pi}}\int_A^\infty x \exp\left(-\tfrac{1}{2}x^2\right)dx = \frac{1}{P\sqrt{2\Pi}}\exp\left(-\tfrac{1}{2}A^2\right)$$

since $B=0$

Given the value of P_{ij}, one can generate the value of A_{ij} and compute the value of $\bar{x}_{ij}$. Since $\mu_W=\mu_{W^*}-A\sigma_W$

$$(40) \qquad \overline{W}_{\cdot j} = \mu_{Wj} + \bar{x}_{ij}\sigma_{Wj} = \mu_{W^*i} + (\bar{x}_{ij} - A_{ij})\sigma_{Wj} = \mu_{W^*} + Z^*_{ij}\sigma_{Wj}$$

Assuming that the standard deviation of the wage offer distribution does not vary among potential wage groups, one can estimate within each income group

$$(41) \qquad \overline{W}_{ij} = a_i + bZ^*_{ij}$$

Again, the constant is an estimate of the mean value of time μ_{W^*i} and the regression coefficient b is an estimate of the standard deviation of the wage offer distribution σ_W.

If one assumes that all women in a given income group share the same price of time ($\sigma_{W^*}=0$), $\overline{W}^*$ equals μ_{W^*} and equation (41) can be used to estimate the housewives' average price of time. If, however, one adopts the other assumption, i.e., that women belonging to the same income-age-education group have identical wage expectations but may differ in the price assigned to their time, $\overline{W}^*$ exceeds μ_{W^*}

$$\overline{W}^* = \mu_{W^*} + \bar{y}\sigma_{W^*} \geq \mu_{W^*} \tag{42}$$

since[18]

$$\bar{y} = E(y/y^* > x) \tag{43}$$
$$= \frac{1}{(1-P)\sqrt{2\Pi}} \exp\left(-\tfrac{1}{2}Z^2\right) > 0$$

One therefore has to compute $\bar{y}$, and using the estimates of μ_{W^*} and σ_{W^*} derived from equation (37) one can obtain an estimate of the housewives' average price of time.

It is worth noting the difference between our two assumptions. According to the first assumption ($\sigma_W = 0$), women who work are those who have the lowest price of time, i.e., are the least productive at home while by the second assumption ($\sigma_{W^*} = 0$), women who work are those who have received the highest wage offers, i.e., are the most productive in the market. This difference carries over to the relationship between the housewives' average value of time and the average market wage rate. According to the first assumption, one expects the housewives' average value of time to exceed the average wage rate, $\overline{W}^* = E(W^* | W^* > \mu_W = \overline{W}) \geq \overline{W}$. By the second assumption $\overline{W}^*$ falls short of $\overline{W}$, i.e. $\overline{W} = E(W | W > \mu_{W^*} = \overline{W}^*) \geq \overline{W}^*$. Actually, it can be shown that these estimates yield extreme limits for the mean value of time.[19]

Finally, if the size of the sample does not allow a very detailed classification of potential wage groups, the number of observations might be too small to allow reliable estimates of equations (37) and (41). In this case the relationship between the mean value of time μ_{W^*} and income I (i.e., $\mu_{W^*i} = g(I_i)$) must be prespecified.

[18] See the Appendix of Gronau (1973).

[19] See the Appendix of Gronau (1973). I have not been able to prove that the first estimate yields an upper limit of μ_{W^*}.

Thus, if, for example, one assumes that there exists a linear relationship

$$\mu_{W^*i} = \alpha_0 + \alpha_1 I_i \tag{44}$$

one can estimate

$$\overline{W}_{ij} = a + b_1 I_i + b_2 Z_{ij} \tag{45}$$
$$\overline{W}_{ij} = a + b_1 I_i + b_3 Z_{ij}^*$$

where $a = \text{est}(\alpha_0)$, $b_1 = \text{est}(\beta_1)$, $b_2 = \text{est}(\sigma_{W^*})$, and $b_3 = \text{est}(\sigma_W)$

IV. The Results

To estimate the value of time of *U.S.* housewives, I used the 1960 Census 1/1000 sample. To isolate the effect of other adults (besides the husband and the wife) on home production and the housewife's value of time I focused on primary families in households without nonrelatives. The sample was restricted to urban white (with no Spanish surname) and Negro married women, spouse present, and consisted of 26,530 observations.

These observations were subclassified according to the woman's race (white, Negro), age (less than 30, 30–49, 40–49, 50+), education (elementary school, high school, college and graduate studies), annual family income when the wife's earnings are excluded (income less than \$2,000, \$2,000–\$2,999, \$3,000–\$3,999, . . . , \$9,000–\$9,999, \$10,000–\$14,999, \$15,000–\$19,999, and \$20,000+) and the existence or nonexistence of children less than three years old. This classification yielded 768 cells (=2x4x4x12x2).

For each cell, I computed the rate of labor force participation (i.e., the percentage of women working or looking actively for work in the week preceding the census), the average income and the average wage of working women. The participation measure reflects the wife's work decision with respect to one specific week

TABLE 1—EDUCATION, AGE, INCOME, AND WAGE DISTRIBUTIONS OF URBAN MARRIED WOMEN BY RACE AND EXISTENCE OF YOUNG CHILD

	Whites			Nonwhites		
	Total[a]	No Young Child[d]	One Young Child	Total[a]	No Young Child	One Young Child
Number in Sample	24,462	19,244	4,079	2,068	1,547	339
Educational Distribution[c]						
Elementary School	24.5	28.1	10.8	42.8	47.7	27.7
High School	58.3	55.4	68.7	48.8	44.2	61.4
College	16.0	15.1	19.3	7.3	7.0	9.4
Graduate Education	1.3	1.3	1.3	1.1	1.0	1.5
Age Distribution[c]						
<30	25.9	14.2	65.9	30.6	17.8	63.1
30–39	27.0	26.5	31.1	29.0	29.6	30.7
40–49	22.1	27.4	2.9	20.6	26.1	5.6
50+	25.1	31.8	0.1	19.9	26.4	0.6
Average Income[b]						
Total Population	6,732	6,877	6,268	3,781	3,822	3,739
Working Women	5,861	5,941	5,070	3,719	3,787	3,216
Average (Potential) Wage						
Working Women	2.019	2.008	2.104	1.528	1.540	1.539
Nonworking Women	2.072	2.083	2.240	1.451	1.329	1.106
Participation Rate						
P	.316	.361	.164	.437	.487	.336
Z	−0.48	−0.36	−0.98	−0.16	−0.03	−0.42

[a] The total includes also mothers with two children younger than three years of age who are not reported in the table.
[b] Wife's earnings excluded.
[c] Given in percent.
[d] Young children are three years of age or less.

and the analysis is therefore limited to the short-run determinants of the housewives' value of time. Likewise, the income measured is current income and does not capture the effect of permanent income on the participation behavior. Finally, the average wage variable is far from being ideal. The 1960 Census reports the number of working hours in the week preceding the census week, the number of weeks worked in 1959, and the wife's earnings in 1959. To obtain a measure of the hourly wage I divided the 1959 earnings by the product of the 1960 weekly hours and the 1959 annual weeks worked and averaged this measure over all women working both in 1959 and in 1960 who belonged to the specific cell. Some of the characteristics of the women belonging to the sample are described in Table 1.

Equation (45) was estimated separately for whites and nonwhites, and within each race group for all women, women without a child younger than three years old and women with one child younger than three.[20] To correct for differences in cell size I fitted a weighted regression, the weights being the number of working women in the cell (i.e., the number of observations used to compute $\overline{W}$). The results of this regression are reported in Table 2.

Of the twelve regressions reported in Table 2, all are significant (at a level of significance of 0.01) except for those relating to white mothers with a child younger

[20] Originally, I distinguished between women with no child less than three years old, mothers of one child less than three, and mothers of two or more children less than three. The last group was, however, too small to allow the estimation of equation (45).

TABLE 2—THE DETERMINANTS OF THE HOUSEWIFE'S VALUE OF TIME

$\mu_{w^*}=\alpha_0+\alpha_1 I$

	Adjusted R^2	Constant a	t	Income[a] I b_1	t	$Z(Z^*)$ b_2	t
			Assumption I ($\sigma_w=0$): $\overline{W}=a+b_1I+b_2Z$				
Whites							
Total	0.22	1.608	21.33	0.8688	7.33	0.2782	2.58
No Young Child[b]	0.22	1.548	20.31	0.8490	7.12	0.2242	2.31
One Young Child	0.02	1.612	5.28	0.9586	1.88	−0.0085	−0.04
Nonwhites							
Total	0.07	1.319	6.46	0.4713	1.01	0.3643	2.81
No Young Child	0.09	1.287	5.97	0.3697	0.77	0.4261	3.37
One Young Child	0.20	1.016	2.96	1.5364	1.63	0.3528	2.34
			Assumption II ($\sigma_{w^*}=0$): $\overline{W}=a+b_1I+b_2Z^*$				
Whites							
Total	0.23	1.073	5.26	0.8348	7.35	0.6548	2.67
No Young Child	0.22	1.146	6.25	0.8198	7.23	0.4994	2.54
One Young Child	0.02	1.632	4.58	0.9615	1.93	−0.0255	−0.33
Nonwhites							
Total	0.08	0.885	3.72	0.3993	0.86	0.5409	3.10
No Young Child	0.09	0.865	3.38	0.3298	0.68	0.5503	3.26
One Young Child	0.20	0.631	1.72	1.3577	1.43	0.4917	2.38

[a] Income is measured in units of $10,000.
[b] Young children are three years of age or less.

than three years old. The coefficients of Z and Z^* are positive and significant for all the remaining ten regressions, but the income coefficient, though positive as expected, is significant only in the case of white women. The standard deviation of the wage offer distribution is found to be in the range of 50–65 cents per hour. The standard deviation of the price of time distribution is 35–43 cents per hour for nonwhites and somewhat lower (22–28 cents per hour) for white women. An increase in the husband's annual earnings of one thousand dollars (roughly 50 cents an hour) increases the mean value of time of his wife by about 8.2–8.7 cents per hour when she is white.

These results are essentially upheld when it is assumed that μ_{W^*} is a linear function of the natural logarithm of income (i.e., $\mu_{W^*}=\alpha_0+\alpha_1 \, log \, (I)$) or when the dispersion of W^* is allowed to vary linearly with income. The assumption that μ_{W^*} is a linear function of $log \, (I)$ resulted in significantly inferior estimates in the case of the white women, and did very little to improve the explanatory power of the regression in the case of nonwhites and thus I do not report the results here.

To estimate the mean price of time in the absence of market opportunities (μ_{W^*}) of the women belonging to a given age-education-income group I computed for each cell

$$\mu_{W^*} = \overline{W} - b_2 Z \tag{46}$$

where it is assumed that $\sigma_W=0$, and

$$\mu_{W^*} = \overline{W} - b_2 Z^* \tag{47}$$

where it is assumed that $\sigma_{W^*}=0$. Weighting each cell by the number of women belonging to the group and summing the estimates over all cells yields an estimate of the mean price of time in the population. Table 3 presents estimates of μ_{W^*} based on the estimates of σ_{W^*} and σ_W (i.e., the coefficient b_2) shown in table 2.

It is found that the mean price of time

TABLE 3—ESTIMATES OF THE MEAN PRICE OF TIME AND THE HOUSEWIVES' AVERAGE PRICE OF TIME BY RACE AND EXISTENCE OF YOUNG CHILDREN

	Whites		Nonwhites		
	Total	No Young Child[a]	Total	No Young Child	One Young Child[a]
Labor Force Participation[b]	31.6	36.1	43.7	48.7	33.6
Average Potential Wage					
All Married Women ($\overline{W}_T$)	2.058	2.058	1.504	1.443	1.415
Working Women ($\overline{W}_W$)	2.019	2.008	1.528	1.540	1.539
Housewives ($\overline{W}_H$)	2.077	2.086	1.492	1.365	1.382
Assumption I ($\sigma_W=0$)					
Mean Price of Time (μ_{W^*})	2.199	2.144	1.544	1.427	1.518
Income Elasticity of $\mu_{W^*}(\epsilon_{W^*I})$	0.27	0.27	0.12	0.10	0.38
Housewives' Price of Time ($\overline{W}^*$)	2.373	2.315	1.822	1.741	1.734
$\overline{W}^*/\overline{W}_H$	1.14	1.11	1.22	1.28	1.26
$\overline{W}^*/\overline{W}_W$	1.18	1.15	1.19	1.13	1.13
Assumption II ($\sigma_{W^*}=0$)					
Mean Price of Time (μ_{W^*})	1.631	1.713	1.071	0.966	1.047
Income Elasticity of $\mu_{W^*}(\epsilon_{W^*I})$	0.34	0.33	0.14	0.13	0.48
Housewives' Price of Time ($\overline{W}^*$)	1.664	1.757	1.100	0.950	1.055
$\overline{W}^*/\overline{W}_H$	0.80	0.84	0.74	0.70	0.76
$\overline{W}^*/\overline{W}_W$	0.82	0.87	0.72	0.62	0.69

[a] Young children are three years of age or less.
[b] Given in percent.

of white married women exceeds that of nonwhites by 40–50 percent (the difference is even larger when one compares white women with no child younger than three years of age with the corresponding nonwhite group). Only part of the difference between these two means can be explained by income differences. The average income of a white family (wife's earnings excluded) exceeds that of a nonwhite family by almost 80 percent ($6,732 vs. $3,781). (See Table 1.) Income seems to have a more substantial effect on the price of time of white women than on the price of time of nonwhites (the exception being nonwhite women with young children). Computing the elasticity of the price of time with respect to income ($\epsilon_{W^*I} = b_1(\bar{I}/\mu_{W^*})$), the estimate in the case of white women is 0.27–0.34 while that of nonwhites is only 0.12–0.14 (see Table 3).[21]

Even if one adopts the highest of these estimates, income can explain at most one-half of the difference between the white and nonwhite mean price of time. To explain the other half, one has to rely on variables whose effect on W^* has not been investigated in this study, such as family composition, and the age and education of the husband and the wife. For example, Table 1 indicates that the nonwhites are more heavily concentrated in the lower age and education groups. If age and education increase the productivity of the woman in the nonmarket as well as in the market sector, this difference should result in a lower mean value of time of nonwhites as compared with white women.

[21] This difference in elasticities can be interpreted as an increase in the sensitivity of the price of time to changes in income as income increases. It is difficult, however, to explain in this way the high value of ϵ_{W^*I} observed in the case of nonwhite mothers of young children.

The attempt to examine the effect of young children on their mothers' value of time was only partly successful. I could not derive meaningful estimates of the parameters of $f(W, W^*)$ for the case of white women with one child younger than three years of age, and the analysis must, therefore, be confined to the nonwhite group.[22]

The difference between the price of time of nonwhite women with one young child and the price of time of women with none is relatively small (6–8 percent). Given the small difference in family income of the two groups (about 2 percent), this difference would have been only slightly affected had we corrected our estimates for income differentials. There is no way of evaluating how this difference would have reacted to an adjustment for age and education. Mothers of young children are, naturally, concentrated in the younger and more educated cohorts. The differences in age composition and education composition tend to offset each other in their effect on the woman's market productivity (the potential wage of the two groups is almost identical). Thus, there is good reason to believe that an adjustment for age and education would not have changed the observed order of magnitude of the child effect.

Assuming $\sigma_{W^*}=0$ the estimate of the average price of time of housewives belonging to a given age-education-income group ($\overline{W}^*$) equals the mean price of time of married women in that group (μ_{W^*}). To obtain an estimate of the average for the whole population one has to compute a weighted average of these estimates, the weights being the number of housewives in each cell. The average $\overline{W}^*$ must not necessarily equal the average value of μ_{W^*} because of different weighting schemes. The estimate of $\overline{W}^*$ usually differs from that of μ_{W^*} insofar as the portion housewives constitute of married women varies from cell to cell.

When it is assumed that $\sigma_W=0$, one has to compute (42) and (43) to obtain an estimate of the housewive's average price of time for each cell. Weighting these estimates by the number of housewives in each cell one obtains an estimate of $\overline{W}^*$ for the whole population. Our estimate of $\overline{W}^*$ are described in Table 3.

Assuming that all women belonging to the same cell expect the same wage rate (i.e., $\sigma_W=0$), it is found that the housewives' average price of time exceeds their average potential wage rate by 14 percent when they are white, and by 22 percent when they are nonwhite. The white nonwhite differentials are explained by the somewhat higher estimate of σ_{W^*} for the nonwhite women, and reflect the different rates of participation resulting in different value of $\bar{y}$. The margins between the price of time $\overline{W}^*$ and the potential wage rate $\overline{W}_H$ are only a little higher if one compares nonwhite housewives with young children with housewives with none.

If one adopts the second assumption (i.e., $\sigma_{W^*}=0$) the housewives' average price of time is expected to fall short of their average potential wage rate. The value of time of white housewives is found to be 80 percent of their potential wage rate while that of nonwhites is 74 percent. The margins for nonwhites with and without children are very similar (70 and 76 percent, respectively).

The preceding comparisons focused on the relationship between the price housewives assign to their own time and what is believed to be their wage expectations. A more ready available basis for comparison is the average wage of working women.

[22] One cannot derive any conclusions from the comparison of the estimate of μ_{W^*} for all whites and that for whites with no young children less than three years of age since the former is not a weighted average of the estimates of μ_{W^*} of whites with young children and μ_{W^*} of whites without young children.

This average wage differs from what is assumed to be the housewives' average potential wage because of the different age and education composition of working and nonworking women.

There is only a very small difference between the average wage of working women ($\overline{W}_W$) and that of housewives ($\overline{W}_H$) both in the case of white and nonwhite women. However, since the former exceeds the latter in the case of nonwhites (by about 2 percent) but the relationship is reversed in the case of whites (the margin being less than 3 percent), the differential between the ratio of $\overline{W}^*/\overline{W}_W$ of whites to nonwhites observed under the assumption $\sigma_W=0$ disappears when one uses as the basis of comparison the average wage of working women ($\overline{W}^*/\overline{W}_W$ being 18–19 percent). On the other hand, these differentials widen if one assumes $\sigma_{W^*}=0$ ($\overline{W}^*/\overline{W}_W$ for whites increasing to 82 percent and $\overline{W}^*/\overline{W}_W$ for nonwhites dropping to 72 percent).

The effect of a change in the basis of comparison is somewhat more pronounced if one compares nonwhites with and without young children. The average wage of a nonwhite working woman with no young child is almost identical with that of a working woman with one young child. The average wage of a working woman exceeds the housewife's potential wage by 11–13 percent. The difference between $\overline{W}^*$ and $\overline{W}_H$ reported in the case of nonwhites (when it as assumed $\sigma_W=0$) is, therefore, cut by one-half when one compares $\overline{W}^*$ with $\overline{W}_W$. On the other hand, the margins increase if one assumes $\sigma_{W^*}=0$, the ratio of $\overline{W}^*/\overline{W}_W$ falling to a level of 0.6–0.7.

V. Some Concluding Remarks

A common practice is to equate the value of time of housewives with that of working women. Given our estimates under two extreme assumptions this practice may involve an error of the magnitude of close to 20 percent in the case of white housewives and an even larger margin of error in the case of nonwhites. One should realize that this margin varies with the housewife's characteristics (for example, her age, education, income, number and age composition of children). Its magnitude depends on the divergence of the wage rate of working women from the potential wage of the nonworking and the relationship between the housewives' average price of time and their expected wage. Thus, it seems that the possible error incurred in the case of nonwhite women is much larger, and may involve an overestimate of $\overline{W}^*$ of up to almost 50 percent.

In another paper (1973), I applied very similar methods to estimate the value of housewives' time in Israel. The labor force participation rate of Israeli married women in 1969 (36 percent) is somewhat higher than the one reported for whites in our sample (32 percent). Assuming $\sigma_W=0$ the estimated coefficient of variation of the price of time distribution (σ_{W^*}/μ_{W^*}) is almost identical for the Israeli and white American women (14 and 13 percent, respectively). Consequently, the ratio of the mean price of time ($\hat{\mu}_W$) and the average potential wage ($\overline{W}_T$) is almost the same in the two groups (1.08 and 1.07, respectively). The income elasticity of W^* is somewhat higher for Israeli women (0.36 vs. 0.27).

The similarity in results is less pronounced if one opts for the assumption $\sigma_{W^*}=0$. The ratio of σ_W to μ_{W^*} is greater for American women than for the Israeli (0.24 vs. 0.14) resulting in a lower $\mu_{W^*}/\overline{W}_T$ ratio for the first group (0.8 vs. 0.9). The estimate of ϵ_{W^*I} is still higher in the Israeli case (0.46 vs. 0.34).

The similarity of some of these results may be comforting but one has to bear in mind the limitations of our model. The assumption that the price of time is unaffected by changes in age and education is

clearly too restrictive. Attempts to remove this assumption according to the lines suggested in my earlier paper have proven to be unsuccessful. As shown by the comparisons of whites and nonwhites, and women with and without young children, the relaxing of this assumption is crucial for the understanding of the determinants of the price of time. The use of disaggregated data may offer a solution to this problem but may involve some other difficulties.

Finally, our procedure is based on the implicit assumption that work in the market and home production do not involve any direct utilities. Even casual observations would indicate that this assumption is wrong. Recognizing the shortcomings of these estimates one has to end this paper with an adequate warning: "Fragile! Handle (the estimates) with care!"

REFERENCES

S. H. Altman and R. J. Barro, "Estimation of a Nursing Labor Supply Function," presented at the Second World Congress of the Econometric Society, Cambridge 1970.

G. S. Becker, "A Theory of the Allocation of Time," *Econ. J.*, Sept. 1965, *75*, 493–517.

Y. Ben-Porath, "Labor Force Participation Rates and the Supply of Labor," disc. pap. no. 206, Harvard Institute of Economic Research 1971.

W. G. Bowen and T. A. Finegan, *The Economics of Labor Force Participation*, Princeton 1969.

G. G. Cain, *Married Women in the Labor Force: An Economic Analysis*, Chicago 1966.

G. R. Ghez and G. S. Becker, "The Allocation of Time and Goods Over the Life Cycle," Center of Mathematical Studies in Business and Economics, Univ. Chicago, rep. 7217, Apr. 1972.

R. Gronau, (1970a) *The Value of Time in Passenger Transportation: The Demand for Air Travel*, occas. pap. 101, Nat. Bur. Econ. Res. New York 1970.

——— (1970b) "An Economic Approach to Marriage: The Intrafamily Allocation of Time," presented at the Second World Congress of the Econometric Society, Cambridge 1970.

——— "The Measurement of Output of the Non-Market Sector—The Evaluation of the Housewives' Time," in *Measurement of Economic and Social Performance*, Nat. Bur. Econ. Res., *Stud. Income and Wealth*, no. 38, forthcoming.

M. Grossman, "On the Concept of Health Capital and the Demand for Health," *J. Polit. Econ.*, Mar./Apr. 1972, *80*, 223–55.

M. B. Johnson, "Travel Time and the Price of Leisure," *Western Econ. J.*, spring 1966 *4*, 135–45.

M. Kosters, "Effects of an Income Tax on Labor Supply," in A. C. Harberger and M. J. Bailey, eds., *The Taxation of Income From Capital*, Washington 1963, 301–24.

K. J. Lancaster, "A New Approach to Consumer Theory," *J. Polit. Econ.*, Apr. 1966, *74*, 132–57.

H. G. Lewis, "On Income and Substitution Effect in Labor Force Participation," mimeo. 1971.

R. Michael, *Education and Consumption Patterns*, Nat. Bur. Econ. Res., forthcoming.

J. Mincer, "Labor Force Participation of Married Women: A Study of Labor Supply," in *Aspects of Labor Economics*, Universities-Nat. Bur. Econ. Res. conference series 15, Princeton 1962, 63–105.

———, "Market Prices, Opportunity Costs and Income Effects," in C. F. Christ, ed., *Measurement in Econometrics: Studies in Mathematical Economics and Econometrics in Memory of Yehuda Grunfeld*, Stanford 1963.

W. C. Mitchell, *The Backward Art of Spending Money and Other Essays*, New York 1950.

H. Ofek, "Efficiency, Substitution and Allocation of Time in Consumption," unpublished mimeo. 1970.

J. P. Smith, "The Life Cycle Allocation of Time in a Family Context," unpublished doctoral dissertation, Univ. Chicago 1972.

Transactions Costs, the Wage Rate, and the Demand for Money

By Dean S. Dutton and William P. Gramm*

Recent work on the demand for money has focused on the productivity of money in yielding services (see Thomas Saving, Jürg Niehans, and Karl Brunner and Allan Meltzer). In the models of Saving and Niehans the use of money reduces transactions time. From such models the traditional results can be derived which emphasize the volume of transactions and the interest rate as determinants of the demand for money. However, if the use of money saves transactions time, it increases the amount of leisure. This suggests an additional determinant of the demand for money, the consumer's valuation of time, i.e., the wage rate. Not only does the demand for money depend on the transacting to be undertaken and the cost of holding money but also on the valuation of the resources which the use of money saves. In equilibrium the marginal valuation of an hour of leisure must be equal to the wage rate. From the assumption of money's usefulness in reducing transactions time it follows that a unit of leisure may be purchased either by reducing work or by increasing money balances. Thus in equilibrium the consumer holds just that level of money balances at which the marginal cost of purchasing an extra unit of leisure via an increase in money holdings would be equal to the wage rate. A change in the wage rate, given the opportunity cost of holding money, results therefore in a change in the relative price of the two ways of purchasing leisure and thereby produces both an income and substitution effect on the demand for money. There is thus a relation between the wage rate and the demand for money independently of changes in income. From our analysis it follows that the demand for money is not determined by but is simultaneously determined with the volume of transactions to be undertaken and the level of income.

The recognition of the leisure time valuation effect of a wage rate change yields both empirical and theoretical implications. For example, the realization of the presence of a leisure time valuation effect in the demand for money offers additional insight into the empirical criteria required to confirm economies of scale in transactions balances. Further the leisure time valuation effect of a change in the wage rate produces a previously neglected effect which alters the macroeconomic adjustment process. Such an effect functions both in the traditional adjustment mechanism where wages and prices are flexible and in the so-called disequilibrium context where traditional adjustment processes do not function. In the disequilibrium model the leisure-price substitution effect provides a theoretical basis for previous assertions about disequilibrium economic behavior and produces adjustment toward equilibrium even in the absence of the traditional adjustment mechanism.

I. The Model

To preserve the dynamic aspects of intertemporal utility maximization and

* Brigham Young University and Texas A&M University, respectively. We are indebted to Thomas Saving for suggestions and criticisms during the various stages of the evolution of this paper. George Borts and an anonymous referee have also made significant contributions to the paper. We also acknowledge the support of a Brigham Young University Research Grant and a Texas A&M Foundation Grant.

yet maintain simplicity, let us assume the consumer behaves so as to maximize

$$(1) \qquad U = U(c, l, a_1)$$

where U is total utility attainment, c is consumption during the period, l is the amount of leisure time taken in the period, and a_1 is the real value of assets held at the end of the period. The assets held at the end of the period form a proxy for future consumption and leisure.

The consumer faces an income constraint (2) and a time constraint (3),

$$(2) \qquad y = wL$$

$$(3) \qquad L + l + T = 1$$

where y is income, w is the wage rate, L is the amount of time spent working, and T is the amount of time spent transacting. Equation (3) states that the three usages of time are mutually exclusive and that their sum exhausts the interval which is arbitrarily defined as 1.

We assume that there are four assets which the consumer can hold; $\bar{m}$, the average stock of money; $\bar{y}$, the average stock of earned income or produced goods; $\bar{c}$, the average stock of consumption goods; and b, bonds which yield rate of interest r. Those assets not held in the form of $\bar{m}$, $\bar{y}$, or $\bar{c}$ are then invested in bonds and earn interest. For simplicity let us assume that income and consumption occur continuously over the time interval and that interest is earned for the portion of the interval during which bonds are held. All arguments of the utility function and the constraints are stated in real terms so that the real value of assets at the end of the period is

$$(4) \quad a_1 = (y-c) + \int_0^1 r(1-t)(y-c)dt + (1+r)a_0 - \sum_i \left(r + \delta_i + \int_0^1 r(1-t)\delta_i dt\right)i$$

where the subscript i identifies $\bar{m}$, $\bar{y}$, or $\bar{c}$ and the variable i measures their quantity; a_0 is equal to initial assets; and δ_i is the real holding cost on the ith stock. The value of assets held at the end of the period is equal to total income earned during the period minus total consumption during the period, plus the interest earnings or payments during the period on the difference $(y-c)$, plus the capitalized value of initial assets, minus the sum of interest lost on the average level of stocks held, minus the storage costs of holding the stocks including the interest lost on the δ_i payments which are assumed to be paid continuously. The time spent in transacting may be formulated as[1]

$$(5) \quad T = T(y, c, \bar{m}, \bar{y}, \bar{c}); \quad T_y > 0,\ T_c > 0,\ T_{\bar{m}},\ T_{\bar{y}},\ T_{\bar{c}} < 0$$

Given w, r, a_0, δ_i the consumer chooses c, l, a_1, y, $\bar{m}$, $\bar{y}$, $\bar{c}$ so as to maximize utility. The function to be maximized is

$$(6) \quad V = U(c, l, a_1) - \lambda_1\left\{a_1 - \left[\left(1+\frac{r}{2}\right)(y-c) + (1+r)a_0 + \sum_i\left(r + \left(1+\frac{r}{2}\right)\delta_i\right)i\right]\right\} - \lambda_2\{y - w[1 - l - T(y, c, \bar{m}, \bar{y}, \bar{c})]\} \quad (i = \bar{m}, \bar{y}, \bar{c})$$

The first-order conditions for the maximization of (6) are

$$(7)\ (a) \quad U_c = \lambda_1\left(1+\frac{r}{2}\right) + \lambda_2 w T_c = \lambda_1\left(1+\frac{r}{2}\right)\frac{[1 + w(T_y + T_c)]}{1 + wT_y}$$

[1] See Saving's analysis of the transactions time function for the general properties of equation (5). As in Saving's analysis, bonds do not enter the transactions time function, and they are assumed to have no holding costs. These assumptions are made for simplicity purposes since we wish to focus attention on money.

$$\text{(b)}\quad U_l = \lambda_2 w = \frac{\lambda_1\left(1+\frac{r}{2}\right)w}{1+wT_y}$$

$$\text{(c)}\quad U_{a_1} = \lambda_1$$

$$\text{(d)}\quad \lambda_1\left(1+\frac{r}{2}\right) = \lambda_2(1+wT_y)$$

$$\text{(e)}\quad -\frac{\lambda_2 wT_{\bar{m}}}{\lambda_1} = -\frac{\left(1+\frac{r}{2}\right)wT_{\bar{m}}}{1+wT_y} = \left[r+\left(1+\frac{r}{2}\right)\delta_{\bar{m}}\right]$$

$$\text{(f)}\quad -\frac{\lambda_2 wT_{\bar{y}}}{\lambda_1} = -\frac{\left(1+\frac{r}{2}\right)wT_{\bar{y}}}{1+wT_y} = \left[r+\left(1+\frac{r}{2}\right)\delta_{\bar{y}}\right]$$

$$\text{(g)}\quad -\frac{\lambda_2 wT_{\bar{c}}}{\lambda_1} = -\frac{\left(1+\frac{r}{2}\right)wT_{\bar{c}}}{1+wT_y} = \left[r+\left(1+\frac{r}{2}\right)\delta_{\bar{c}}\right]$$

The terms on the right-hand side of (7a) and (7b) are equal to λ_1 times the relative opportunity costs of consumption and leisure, respectively. Equation (7a) specifies that the marginal utility of consumption must equal its opportunity cost in terms of the marginal utility of asset acquisition foregone and the transactions time cost incurred valued at the wage rate. Equation (7b) requires that the marginal utility of leisure equal the wage rate valued in terms of the marginal utility of end of the period assets. Equation (7c) defines λ_1 as the marginal utility of end of the period assets, and (7d) yields the relation between λ_1 and λ_2. Equations (7e), (7f), and (7g) require that the marginal returns to holding average stocks of money balances, goods produced and goods consumed equal the respective opportunity cost of holding the noninterest bearing stocks in terms of the interest earnings foregone and holding costs incurred. The term $(1+wT_y)$ in denominators of a, b, e, f, and g reflects the fact that the transactions time saved cannot be entirely converted into income, given fixed stocks, because increased income requires increased transactions time.

Equations (7) and (4) plus (3) and (5) substituted into (2) make up nine equations in nine unknowns with six parameters. Solving for the optimum money stock $\bar{m}$, the demand for money can be written as

$$\text{(8)}\qquad \bar{m} = m(w, r, a_0, \delta_{\bar{m}}, \delta_{\bar{y}}, \delta_{\bar{c}})$$

$$\text{(9)}\quad A\begin{bmatrix} dc \\ dl \\ da_1 \\ dy \\ di \\ d\lambda_1 \\ d\lambda_2 \end{bmatrix} = \begin{bmatrix} \frac{\lambda_1}{2} & \lambda_2 T_c & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -\frac{\lambda_1}{2} & \lambda_2 T_y & 0 & 0 \\ \lambda_1[1+\delta_j/2] & \lambda_2 T_j & 0 & \lambda_1[1+r/2] \\ -[(y-c)/2+a_0-\sum_i(1+\delta_i/2)i] & 0 & -(1+r) & (1+r/2)j \\ 0 & [1-l-T(y,c,i)] & 0 & 0 \end{bmatrix}\begin{bmatrix} dr \\ dw \\ da_0 \\ d\delta_j \end{bmatrix}$$

$$
\text{(9a)}\quad
\begin{bmatrix}
U_{cc}-\lambda_2 wT_{cc} & U_{cl} & U_{ca_1} & -\lambda_2 wT_{cy} & -\lambda_2 T_{ci} & -(1+r/2) & -wT_c \\
U_{lc} & U_{ll} & U_{la_1} & 0 & 0 & 0 & -w \\
U_{a_1c} & U_{a_1l} & U_{a_1a_1} & 0 & 0 & -1 & 0 \\
-\lambda_2 wT_{yc} & 0 & 0 & -\lambda_2 wT_{yy} & -\lambda_2 wT_{yi} & (1+r/2) & -(1+wT_y) \\
-\lambda_2 wT_{jc} & 0 & 0 & -\lambda_2 wT_{jy} & -\lambda_2 wT_{ji} & -[r+(1+r/2)\delta_j] & -wT_j \\
-(1+r/2) & 0 & -1 & (1+r/2) & -[r+(1+r/2)\delta_i] & 0 & 0 \\
-wT_c & -w & 0 & -(1+wT_y) & -wT_i & 0 & 0
\end{bmatrix}
$$

$$
\text{(10)}\quad \text{(a)}\quad \frac{\partial \bar{m}}{\partial w} = \frac{\lambda_2 T_c D_{c\bar{m}}}{D} + \frac{\lambda_2 D_{l\bar{m}}}{D} + \frac{\lambda_2 T_y D_{y\bar{m}}}{D} + \frac{\lambda_2 T_{\bar{m}} D_{\bar{m}\bar{m}}}{D} + \frac{\lambda_2 T_{\bar{y}} D_{\bar{y}\bar{m}}}{D} + \frac{\lambda_2 T_{\bar{c}} D_{\bar{c}\bar{m}}}{D} + \frac{[1 - l - T(y, c, \bar{m}, \bar{y}, \bar{c})] D_{\lambda_2 \bar{m}}}{D}
$$

$$
\text{(b)}\quad \frac{\partial \bar{m}}{\partial a_0} = \frac{-(1+r) D_{\lambda_1 \bar{m}}}{D}
$$

$$
\text{(c)}\quad \frac{\partial \bar{m}}{\partial r} = \frac{\frac{1}{2}\lambda_1 D_{c\bar{m}}}{D} - \frac{\frac{1}{2}\lambda_1 D_{y\bar{m}}}{D} + \frac{\lambda_1[1 + \delta_{\bar{m}}/2] D_{\bar{m}\bar{m}}}{D} + \frac{\lambda_1[1 + \delta_{\bar{y}}/2] D_{\bar{y}\bar{m}}}{D} + \frac{\lambda_1[1 + \delta_c/2] D_{\bar{c}\bar{m}}}{D} - \frac{[(y-c)/2 + a_0 - \sum_i (1 + \delta_i/2) i] D_{\lambda_1 \bar{m}}}{D}
$$

$$
\text{(d)}\quad \frac{\partial \bar{m}}{\partial \delta_j} = \frac{\lambda_1(1 + r/2) D_{j\bar{m}}}{D} + \frac{(1 + r/2) j D_{\lambda_1 \bar{m}}}{D} \qquad (j = \bar{m}, \bar{y}, \bar{c})
$$

The derivatives of (8) can be derived by totally differentiating the system of equations (7), (4), and (2) after substituting in (3) and (5). The differentiation yields the system shown as (9), where A is the 9x9 matrix shown as (9a), and $j,\ i = \bar{m},\ \bar{y},\ \bar{c}$. Solving for the changes in $\bar{m}$ with respect to changes in the parameters, we have the system shown as (10), where D is the determinant of A in (9a) which is assumed to be negative definite as a condition for maximization, and D_{ij} is the cofactor of the ith row and jth column.

In order to specify the nature of the derivatives of the demand for money function both to clarify the source of the various disturbances and to sign the derivates, it is helpful to associate the terms of the equations of (10) with effects about which we can make some assumptions. In this frame of reference, let us gauge the change in the choice variables with respect to a change in $\delta_{\bar{m}}$ the real cost of holding money. Solving for the change in the kth decision variable with respect to $\delta_{\bar{m}}$

$$
\text{(11)}\quad \frac{\partial k}{\partial \delta_{\bar{m}}} = \frac{\lambda_1(1 + r/2) D_{\bar{m}k}}{D} + \frac{(1 + r/2)\bar{m} D_{\lambda_1 k}}{D} \qquad (k = c, l, a_1, y, \bar{m}, \bar{y}, \bar{c})
$$

Note further that

$$\frac{\partial k}{\partial[(1+r)a_0]} = -\frac{D_{\lambda_1 k}}{D} \tag{12}$$

where $(1+r)a_0$ is the end of period capitalized value of the stock of initial real assets. Thus a change in the holding cost of money produces an income and substitution effect on the decision variables. The income term is the capitalized compensation required to maintain the consumer at the same level of utility attainment after $\delta_{\bar{m}}$ changes.

The change in transactions time with respect to a change in the holding cost of money balances may be stated in utility constant form as

$$\left(\frac{\partial T}{\partial \delta_{\bar{m}}}\right)_{\bar{u}} = T_y\left(\frac{\partial y}{\partial \delta_{\bar{m}}}\right)_{\bar{u}} + T_c\left(\frac{\partial c}{\partial \delta_{\bar{m}}}\right)_{\bar{u}} + T_{\bar{m}}\left(\frac{\partial \bar{m}}{\partial \delta_{\bar{m}}}\right)_{\bar{u}} + T_{\bar{y}}\left(\frac{\partial \bar{y}}{\partial \delta_{\bar{m}}}\right)_{\bar{u}} + T_{\bar{c}}\left(\frac{\partial \bar{c}}{\partial \delta_{\bar{m}}}\right)_{\bar{u}} \tag{13}$$

Noting from (7d) that $\lambda_1 = \lambda_2(1+wT_y)/(1+r/2)$ and inserting the substitution effect terms from (11) into (13)

$$\left(\frac{\partial T}{\partial \delta_{\bar{m}}}\right)_{\bar{u}} = (1+wT_y)\left(\frac{T_y\lambda_2 D_{\bar{m}y}}{D} + \frac{T_c\lambda_2 D_{\bar{m}c}}{D} + \frac{T_{\bar{m}}\lambda_2 D_{\bar{m}\bar{m}}}{D} + \frac{T_{\bar{c}}\lambda_2 D_{\bar{c}\bar{m}}}{D} + \frac{T_y\lambda_2 D_{\bar{y}\bar{m}}}{D}\right) \tag{14}$$

The last term of (10a) is an income effect term. This may be seen by adding a shift parameter π to the income function in the second constraint of (6) and solving for $\partial\bar{m}/\partial\pi = D_{\lambda_2\bar{m}}/D$. Using the results of (14) and substituting from (13), (11, $k=l$) and noting the presence of the income effect term produced by a parallel shift in the constraint, (10a) can be written as

$$\frac{\partial \bar{m}}{\partial w} = \frac{1}{(1+wT_y)}\left\{\left(\frac{\partial T}{\partial \delta_{\bar{m}}}\right)_{\bar{u}} + \left(\frac{\partial l}{\partial \delta_{\bar{m}}}\right)_{\bar{u}}\right\} + [1 - l - T(y, c, \bar{m}, \bar{y}, \bar{c})]\frac{\partial \bar{m}}{\partial \pi} \tag{15}$$

which expresses the derivative of the demand for money with respect to the wage rate in terms of the weighted sum of three other derivatives. The sum $(\partial T/\partial\delta_{\bar{m}})_{\bar{u}} + (\partial l/\partial\delta_{\bar{m}})_{\bar{u}}$ is proportional to the substitution effect of a change in the wage rate and $\partial\bar{m}/\partial\pi$ is proportional to the income effect. Since the weights are positive it remains only to sign the two terms $(\partial T/\partial\delta_{\bar{m}})_{\bar{u}} + (\partial l/\partial\delta_{\bar{m}})_{\bar{u}}$ and $\partial\bar{m}/\partial\pi$ in order to determine the sign of $\partial\bar{m}/\partial w$.

Assuming the normality of goods and inputs $\partial\bar{m}/\partial\pi > 0$. The increased demand for both consumption and leisure as income or assets increase creates an increase in the derived demand for money through the transactions time function. First, in order to satisfy the increased demand for leisure the consumer will reduce work and reduce transactions time by holding larger money balances. Second, in order to consume more—which affects transactions time positively—without enjoying less leisure the consumer will work less and hold larger money balances to offset the effect of increased consumption on transactions time.

Holding utility constant, an increase in $\delta_{\bar{m}}$ which increases the cost of purchasing leisure via money holdings, will induce the consumer to reduce his consumption of leisure and increase desired consumption of goods and holdings of end of period assets. Transactions time will increase first, because money, one input in the transactions time function, has become relatively more expensive (recalling that $\partial T/\partial\bar{m} < 0$), and second because consumption has increased (recalling that

$\partial T/\partial c>0$). Therefore $(\partial T/\partial \delta_{\bar{m}})_{\bar{u}}>0$ and $(\partial l/\partial \delta_{\bar{m}})_{\bar{u}}<0$. Since the increase in the holding cost of money increases the cost of obtaining leisure by reducing transactions time relative to the cost of obtaining leisure via reduction in work time, there will be a substitution by the consumer of work time for transactions time in purchasing his desired level of leisure. That is, because of the change in the relative price the consumer will reduce work time and increase transactions time. Therefore $(\partial L/\partial \delta_{\bar{m}})_{\bar{u}}<0$. Noting that $(\partial T/\partial \delta_{\bar{m}})_{\bar{u}}+(\partial l/\partial \delta_{\bar{m}})_{\bar{u}}+(\partial L/\partial \delta_{\bar{m}})_{\bar{u}}=0$, it then follows that $[(\partial T/\partial \delta_{\bar{m}})_{\bar{u}}+(\partial l/\partial \delta_{\bar{m}})_{\bar{u}}]>0$ and consequently $\partial \bar{m}/\partial w>0$. Given our assumptions, then, it follows that both the substitution effect and the income effect of a change in the wage rate are positive.

The normality of the arguments of the utility function and the transactions time function makes $\partial \bar{m}/\partial a_0>0$. A rise in assets causes a rise in consumption of goods but by normality the consumer is loath to decrease leisure so he holds larger money balances and works less. A rise in assets will thus cause a rise in money holding.

The terms of the derivative of money balances with respect to the interest rate can be identified and partially signed by substituting the relations derived by transformation (11) and formulation (10b) into (10c). After such substitution

$$(16)\quad \frac{\partial \bar{m}}{\partial r} = \left\{\left(\frac{2+\delta_{\bar{m}}}{2+r}\right)\left(\frac{\partial \bar{m}}{\partial \delta_{\bar{m}}}\right)_{\bar{u}} + \left(\frac{2+\delta_{\bar{y}}}{2+r}\right)\left(\frac{\partial \bar{y}}{\partial \delta_{\bar{m}}}\right)_{\bar{u}} + \left(\frac{2+\delta_{\bar{c}}}{2+r}\right)\left(\frac{\partial \bar{c}}{\partial \delta_{\bar{m}}}\right)_{\bar{u}}\right\} + \frac{1}{2+r}\left(\frac{\partial c}{\partial \delta_{\bar{m}}} - \frac{\partial y}{\partial \delta_{\bar{m}}}\right)_{\bar{u}} + \frac{1}{(1+r)}\left[\frac{(y-c)}{2} + a_0 - \sum_i (1+\delta_i/2)i\right]\frac{\partial \bar{m}}{\partial a_0}$$

The first three terms of (16) are the substitution terms produced on the stocks held by a change in the cost of holding money. Assuming own price dominance of the three effects will not assure the dominance of the first term since each term is weighed by its holding costs. The value of $(\partial \bar{m}/\partial \delta_{\bar{m}})_{\bar{u}}$ is negative, and the values $(\partial \bar{y}/\partial \delta_{\bar{m}})_{\bar{u}}$ and $(\partial \bar{c}/\partial \delta_{\bar{m}})_{\bar{u}}$ are positive or negative depending on whether $\bar{m}$ and $\bar{y}$, $\bar{c}$ are substitutes or complements. The second substitution effect is the effect of a change in $\delta_{\bar{m}}$ on consumption and income. A rise in $\delta_{\bar{m}}$ makes transacting more expensive and reduces both purchases (c) and sales (y) made by the consumer. If the impact of a rise in $\delta_{\bar{m}}$ is symmetric, the substitution effect of a change in $\delta_{\bar{m}}$ is the same on both c and y so that the second substitution term is zero. Assuming the normality of money holding $\partial \bar{m}/\partial a_0>0$, while the sign of $(y-c/2)+a_0-\sum_i (1+\delta_i/2)i$ (the derivative of the a_1 with respect to r), is positive, negative, or zero depending on the relative magnitudes of $(y-c)$, a_0, $\bar{m}$, $\bar{y}$, $\bar{c}$, $\delta_{\bar{m}}$, $\delta_{\bar{y}}$, and $\delta_{\bar{c}}$.

There is thus a negative own price substitution effect, indeterminate cross price substitution effects, indeterminate trade effect, and indeterminate income effect produced on the demand for money when the interest rate changes. The own price substitution effect of a rise in r makes money more expensive to hold and causes $\bar{m}$ to fall. The cross price substitution effects alter the holding of stocks of substitutes and complements. The trade effect alters sales and purchases and its sign depends on whether c or y is more affected by a change in $\delta_{\bar{m}}$. How a rise in r affects wealth and by normality affects the demand for money depends on the composition of a_1 and the sign of $\{[(y-c)/2]+a_0\}$. A rise in the interest rate raises the holding

cost of the noninterest bearing stocks held and lowers wealth. The impact of a rise in the interest rate on $\{[(y-c)/2]+a_0\}$ depends on whether the individual is a net creditor or debtor. If he is a net creditor (debtor) a rise in the interest rate raises (lowers) his wealth. Therefore, for the net debtor the income effect of a change in the interest rate is negative, and for the net creditor it may be positive, negative, or zero depending on the composition of his portfolio.[2]

A change in the holding costs of assets produces both an income and substitution effect on the demand for money as shown in (10d). As assumed in (10b), $-D_{\lambda_1\bar{m}}/D > 0$ so that the income term of (10d) is negative. This can be seen by noting that a rise in the price of a stock increases the price of leisure and lowers wealth. In the own price case where $i=\bar{m}$ the substitution term must be negative as a condition for maximization. When $i=\bar{y}$, $\bar{c}$ the sign of the substitution term of (10d) depends on the assumed relation between $\bar{m}$ and $\bar{y}$ or $\bar{c}$ in the transactions costs function.[3]

II. An Empirical Note

The demand function for money generated by the above analysis (8) states the demand for real money balances as a function of the real wage, the interest rate, the real value of assets and real storage costs on goods produced, goods consumed, and money. Since suitable data are not available for storage costs, the empirical specification of (8) is limited to the real wage (w), the interest rate (r), and the real value of initial assets (a_0). Assuming away distribution effects we may aggregate over individuals' demands for money. Such aggregation eliminates the income effect of a change in the interest rate on internal debt.[4] Since the theoretical analysis above is cast in terms of a general equilibrium system the standard assumption is made regarding the equality of the real money stock and the demand for real balances. The logic of such an assumption is that while society does not determine the nominal stock of money it holds, it does determine the price level and thus the real money stock. Further, the ratio of the money stock held by consumers to the total stock is assumed fixed so that data on the total money stock can be employed in the tests below.

The demand function to be estimated is assumed to be linear in the *logs* of its arguments so that the equation to be estimated is

$$(17)\quad ln\, \bar{m}_t = ln\, \beta_0 + \beta_1\, ln\, w_t + \beta_2\, ln\, r_t + \beta_3\, ln\, a_t + u_t$$

where u_t is assumed $\sim N(0, \sigma^2)$. In estimates of equation (17), three different definitions of the interest rate and four different definitions of the real value of assets were employed. Short-, intermediate- and long-term interest rates were defined

[2] The above analysis follows only if debts are recontracted each time the interest rate changes. Given recontracting holders of bonds as net creditors incur a rise (fall) in wealth as the interest rate rises (falls). If debts were not recontracted the market value of existing bonds would fall (rise) as the interest rate rises (falls).

[3] It is easily shown that the cross price substitution effect depends on the production relation in the transaction time function between $\bar{m}$ and $\bar{y}$, $\bar{c}$. Note for example that

$$\frac{\partial \bar{m}}{\partial \delta_{\bar{y}}} = \frac{\lambda_1(1+r/2)D_{\bar{y}\bar{m}}}{D} + \frac{(1+r/2)\bar{y}D_{\lambda_1\bar{m}}}{D}$$

so that

$$\frac{\lambda_1(1+r/2)D_{\bar{y}\bar{m}}}{D} = \left(\frac{\partial \bar{m}}{\partial \delta_{\bar{y}}}\right)_{\bar{u}}$$

which is positive if $\bar{y}$ and $\bar{m}$ are substitutes and negative if they are complements.

[4] Public debt can be treated as internal debt if individuals take account of the impact of the burden of interest payments on the public debt on future taxes or the flow of government services. In the empirical analysis below, government debt is not counted as part of national wealth so that the expected sign of the income term of (10c) is negative.

as follows:

r_1 (short term) = rate on prime commercial paper (4–6 months),

r_2 (intermediate) = rate on short-term business loans,

r_3 (long term) = Moody's AAA bond yields.

Four definitions of the real value of assets were generated by assuming that the flow of services yielded by government assets are valued by the consumer either at par with the flow of services from private assets or at zero. Demand deposits net of reserves were assumed to be either net additions to total wealth or pure inside money which is the debt of the issuer. Thus

$$(18)\qquad a_1 = \frac{A_T - M_{gS} + C_P + D_d}{P}$$

$$a_2 = \frac{A_T - M_{gS} + M_b}{P}$$

$$a_3 = \frac{A_T - A_g + C_P + D_d}{P}$$

$$a_4 = \frac{A_T - A_g + M_b}{P}$$

where A_T is total nominal nonmonetary wealth, M_{gS} is the nominal value of monetary gold and silver held by the government, C_P is the nominal value of currency held by the public, D_d is the nominal value of demand deposits held by the public, M_b is the nominal monetary base, A_g is the nominal value of government assets, and P is the price level.

Initial estimates of (17) using r_i and a_j yielded Durbin-Watson statistics which indicated the presence of significant positive autocorrelation. Employing a two-step Orcutt procedure (see Donald Cochrane and Guy Orcutt) adjustments were made to eliminate the autocorrelation in the residuals.[5] Table 1 presents the estimates of equation (17) after adjustments have been made for autocorrelation, and the preadjustment Durbin-Watson statistics.

The most consistent result found in Table 1 is the statistical significance of w in all equations estimated. The coefficient of the wage rate is positive and statistically significant at the 1 percent level or better in every equation, regardless of the definition of the interest rate and real assets employed. Since equation (17) is estimated in the *logs* of its arguments, the coefficients of w in Table 1 are the wage rate elasticities of the demand for money. The elasticities vary from 1.27 when r_1 and a_2 are the assumed interest rate and asset level definitions to .76 when r_3, a_1 or r_3, a_3 are employed in the test equation.

The sign of r_i is negative in all equations and statistically significant at the 1 percent level or better in four equations, in all of which r_3, the long-term rate, was employed. The coefficient of the interest rate is statistically significant at the 5 percent level when the intermediate-term rate is employed. When the short-term rate is employed the coefficient of r is statistically significant at the 5 percent level in one equation and at the 10 percent level in the other three equations. The superiority of the long-term rate in producing a significant coefficient for r is compatible with other studies of the demand for money which have used the long-term rate

[5] Assuming a first-order autoregressive disturbance structure

$$u_t = \rho u_{t-1} + \epsilon_t$$

where $\epsilon_t \sim N(0, \sigma^2)$, $\hat{\rho}$ is gauged as follows

$$\hat{\rho} = \frac{\sum_{t=2}^{N} \hat{u}_t \hat{u}_{t-1}}{\sum_{t=1}^{N} \hat{u}_t^2} \left(\frac{N-1}{N}\right)$$

Equation (17) is reformulated as

$$\begin{aligned} \ln \bar{m}_t - \hat{\rho} \ln \bar{m}_{t-1} = \ln \beta_0(1-\hat{\rho}) &+ \beta_1(\ln w_t - \hat{\rho} \ln w_{t-1}) \\ &+ \beta_2(\ln r_{it} - \hat{\rho} \ln r_{it-1}) \\ &+ \beta_3(\ln a_{jt} - \hat{\rho} \ln a_{jt-1}) + \epsilon_t \end{aligned}$$

TABLE 1—ESTIMATES OF EQUATION (17) EMPLOYING THREE DIFFERENT DEFINITIONS OF r AND FOUR DIFFERENT DEFINITIONS OF a USING *U.S.* ANNUAL DATA FOR THE 1919–58 PERIOD

r, a Identification	Coefficients of Constant	w	r	a	R^2	d
r_1, a_2	0.46	+1.11 (0.30)[a]	−0.07 (0.04)[b]	−0.40 (0.25)[c]	.800	0.43
r_2, a_1	0.18	+0.85 (0.28)[a]	−0.21 (0.11)[b]	−0.45 (0.22)[b]	.607	0.30
r_3, a_1	0.17	+0.76 (0.27)[a]	−0.49 (0.16)[a]	+0.35 (0.20)[b]	.604	0.30
r_1, a_2	0.61	+1.27 (0.29)[a]	−0.06 (0.04)[c]	+0.26 (0.25)	.800	.057
r_2, a_2	0.28	+1.00 (0.27)[a]	−0.22 (0.10)[b]	+0.34 (0.22)[b]	.642	0.36
r_3, a_2	0.23	+0.86 (0.26)[a]	−0.51 (0.16)[a]	+0.27 (0.20)[c]	.620	0.31
r_1, a_3	0.27	+0.96 (0.28)[a]	−0.07 (0.05)[c]	+0.46 (0.23)[b]	.700	0.40
r_2, a_3	0.17	+0.82 (0.28)[a]	−0.21 (0.11)[b]	+0.46 (0.21)[b]	.581	0.29
r_2, a_3	0.17	+0.76 (0.26)[a]	−0.49 (0.16)[a]	+0.36 (0.20)[b]	.605	.031
r_1, a_4	0.59	+1.26 (0.28)[a]	−0.07 (0.05)[c]	+0.26 (0.24)	.791	0.55
r_2, a_4	0.27	+0.99 (0.27)[a]	−0.22 (0.11)[b]	+0.34 (0.21)[c]	.617	0.33
r_3, a_4	0.23	+0.87 (0.26)[a]	−0.52 (0.16)[a]	+0.27 (0.20)[c]	.621	0.31

[a] Statistically significant at the 1 percent level.
[b] Statistically significant at the 5 percent level.
[c] Statistically significant at the 10 percent level.

Note: The d reported is for the original equation before corrections were made for the presence of autocorrelation. Standard errors are in parenthesis below coefficients.

Sources: Data for M, P, r_1, r_2, r_3, and w from 1919–57 were taken from U.S. Bureau of the Census, pp. 646, 125–126, 654, 655, 656, and 692, respectively. The various asset variables were calculated using data from R. W. Goldsmith (1955, 1962).

(Meltzer, p. 140), (Gregory Chow, p. 128). Further the highest interest elasticities are obtained when the maturity length of the interest rate is increased. However, the interest elasticities of r_3 are smaller than those obtained by Meltzer using traditional formulations of the demand for money function (pp. 134, 139, 143).

The various definitions of the real value of asset holdings have the expected positive sign in all equations and the coefficient of a is statistically significant at

the 5 percent level in six equations and at the 10 percent level in four equations. Since the wage rate has an income term as formulated in equation (11a), the elasticity of the asset variable would be expected to be lower than in an equation where only r and a are employed in the test equation. Meltzer, employing only r and a for the 1900–58 period, found the asset elasticity of the demand for real money balances to be 1.11 (p. 134) as compared with the .46 to .26 coefficients shown in Table 1.[6]

III. Theoretical and Empirical Implications in Disequilibrium States

The leisure time valuation or substitution effect of a change in the wage rate on the demand for money has important implications with regard to the adjustment process both in the traditional cases where wages and prices are flexible, and instantaneous adjustment is assumed, and in so-called disequilibrium systems where wages and prices are partially invariant and markets do not adjust. It is in the latter cases that the impact of the leisure time valuation or substitution effect on the demand for money appears most novel.

The leisure time valuation effect on the demand for money is very interesting when viewed within the context of the disequilibrium analysis of Robert Barro and Herschel Grossman. From their model, it follows that a reduction in commodity demand and output produces a decline in the demand for labor with a corresponding reduction in employment. Even if wages do fall in response to the resulting excess supply of labor, they conclude that the only result is to induce man-hours of labor to leave the labor force, not to increase the quantity of man-hours of labor demanded. Our analysis, when applied to this model, supplies a force which acts to restore equilibrium in both the commodity and labor markets. Take the Barro-Grossman example of excess labor supply resulting from excess supply in the commodities market (pp. 88–90). Consumers find themselves oversupplied with leisure due to involuntary unemployment and as a result of money's role in providing leisure there is a reduction in the demand for money. To the involuntarily unemployed the wage rate is no longer the marginal cost of leisure. To the extent that wages fall in response to the excess supply of labor there is a further reduction in the demand for money by those who are still employed. This excess supply of money tends to increase commodity demand and thus labor demand, employment, and output.

In the Barro-Grossman case of excess commodity demand (pp. 90–92) our analysis gives their results a choice theoretic foundation and extends them somewhat. They postulate that consumers may react to frustrated commodity demand in two ways: first, by involuntarily augmenting money balances or second, by reducing labor supply. They conclude that generally some combination of the two options will always be optimal. According to our analysis, frustrated commodity demand would be assumed to mean frustrated consumer searches and therefore a shift in the transactions time function. In terms of the first-order conditions (7) this would mean an increase in the price of consumption through an increased T_c. Since the price of consumption relative to leisure has risen, consumers will want to consume less commodities and more leisure directly by reducing work. The fall in work time, given

[6] Meltzer employs a definition of wealth which differs from the four employed in the test above. He excludes all government assets but includes all government debt. This is equivalent to assuming that the consumer valuation of the flow of services from government assets is zero and that consumers do not respond to changes in the real tax burden produced by changes in the level of interest-bearing government debt.

the wage rate, will reduce income, y, and, therefore, reduce T_y in the denominator of (7e), (7f), (7g). The equilibrium condition (7e) is again satisfied by increasing money balances which reduces $T_{\bar{m}}$. This yields a theoretical basis for the Barro-Grosman results of reduced labor supply and increased demand for money.[7]

In the presence of involuntary unemployment the wage rate overstates the opportunity cost of time and the presence of involuntary unemployment would be expected to distort the relation between the wage rate and the demand for money. It is, therefore, useful to reformulate the empirical specification of the demand for money, entering the unemployment rate (γ) into the demand for money function as an explanatory variable. Since for the unemployed the valuation of time (V_T) is less than the market wage rate, it follows that $[(\partial\bar{m}/\partial V_T)(\partial V_T/\partial\gamma)]<0$. Adding the unemployment rate as an explanatory variable of (17) and employing the definitions of r and a used in Section II, the results of the respecification of the demand for money equation are shown in Table 2.

The estimates of Table 2 are those obtained after the Cochrane-Orcutt Interative Technique for eliminating autocorrelation was employed. Since estimates of ρ were approximately 1, the constant term had to be suppressed. While the absence of a constant term tends to bias the estimate of R^2 downward, in every equation employing γ the R^2 is .99. The overall fit is, therefore, improved by introducing the disequilibrium phenomenon into the test equation. Further the coefficient of the unemployment rate is negative and statistically significient at the 1 percent level in every equation. The coefficient of the wage rate is positive and statistically significant at the 1 percent level when the short-term interest rate is employed and at the 5 percent level when the intermediate-term rate is employed. The coefficient of the level of assets is not statistically significant at the 10 percent level in any equation and in many cases it has the wrong sign.[8]

The results of Table 2 suggest that the disequilibrium effect of unemployment had the expected negative effect on the demand for money. To gauge the potential magnitude of the leisure time valuation effect of unemployment, we may employ the .07 estimate of the elasticity of demand for money with respect to the unemployment rate to estimate the decline in money holding in response to unemployment's effect

[7] The analysis can be extended further. The increase in the demand for money caused by the frustrated commodity demand tends to eliminate the excess demand in the commodity market. On the other hand, however, the decreased supply of labor tends to reduce aggregate supply and thus tends to offset the reduction in excess commodity demand caused by increased demand for money. There is still another effect, however. The reduction in the supply of labor puts upward pressure on wages, and increased wages increase the demand for money which again tends to reduce the excess demand in the commodity market. There are thus forces which tend to both reduce and accentuate the initial excess commodity demand. Our observations of a stable system would lead us to believe, however, that the forces predominate which work to eliminate excess commodity demand.

[8] It is interesting to compare the results obtained by employing income instead of the wage rate in the estimate of the demand for money function. From Table 2, with definitions r_1, a_1:

$$\begin{array}{llllll} \bar{m}= & 0.29\,w & -0.11\,r & +0.08\,a & -0.07\,\gamma & (R^2=.99) \\ & (0.09)\dagger & (0.04)\dagger & (0.24) & (0.02)\dagger & \end{array}$$

Employing y in place of w

$$\begin{array}{llllll} \bar{m}= & 0.41\,y & -0.14\,r & -0.39\,a & -0.03\,\gamma & (R^2=.99) \\ & (0.18)^* & (0.04)\dagger & (0.12) & (0.03) & \end{array}$$

The dagger indicates significance at the 1 percent level, and the asterisk at 5 percent. Though the coefficient of y is positive and statistically significant at the 5 percent level, the elasticity of the demand for money with respect to y is larger than the elasticity of the demand for money with respect to w. Further, the coefficient of γ in the equation employing y is not statistically significant at the 10 percent level. This result could reflect the fact that measured y is encompassing some of the impact of employment conditions so that demand for money functions fitted employing y yield good fits because y yields the impact of w after leisure choice decisions have been made and also picks up the impact of employment conditions. This supposition is borne out by the fact that replacing w with y in the general equilibrium formulation of (17) raises the R^2 of the equation from .80 to .99.

TABLE 2—ESTIMATES OF EQUATION (17) AS ADAPTED TO ACCOUNT FOR THE DISEQUILIBRIUM IMPACT OF UNEMPLOYMENT ON THE EFFECTIVE OPPORTUNITY COST OF TIME EMPLOYING THREE DIFFERENT DEFINITIONS OF r AND FOUR DIFFERENT DEFINITIONS OF a USING U.S. ANNUAL DATA FOR THE 1919–58 PERIOD

r, a Identification	Coefficients of w	r	a	γ	R^2	d
r_1, a_1	0.29 (0.09)[a]	−0.11 (0.04)[a]	0.08 (0.24)	−0.07 (0.02)[a]	.988	.00
r_2, a_1	0.23 (0.10)[b]	−0.30 (0.10)[a]	0.03 (0.24)	−0.07 (0.02)[a]	.988	.00
r_3, a_1	0.00 (0.11)	−0.67 (0.15)[a]	−0.19 (0.23)	−0.06 (0.02)[a]	.990	.00
r_1, a_2	0.27 (0.09)[a]	−0.12 (0.04)[a]	0.01 (0.23)	−0.08 (0.02)[a]	.988	.00
r_2, a_2	0.19 (0.10)[b]	−0.31 (0.10)[a]	−0.09 (0.24)	−0.07 (0.02)[a]	.988	.00
r_3, a_2	−0.03 (0.10)	−0.69 (0.15)[a]	−0.27 (0.23)	−0.07 (0.02)[a]	.991	.00
r_1, a_3	0.30 (0.09)[a]	−0.11 (0.04)[a]	0.11 (0.23)	−0.07 (0.02)[a]	.988	.00
r_2, a_3	0.24 (0.10)[b]	−0.29 (0.10)[a]	0.06 (0.23)	−0.06 (0.02)[a]	.988	.00
r_3, a_3	0.02 (0.11)	−0.66 (0.15)[a]	−0.16 (0.22)	−0.06 (0.02)[a]	.990	.00
r_1, a_4	0.28 (0.09)[a]	−0.12 (0.04)[a]	−(0.00) (0.17)	−0.07 (0.02)[a]	.988	.00
r_2, a_4	0.19 (0.09)[b]	−0.31 (0.10)[a]	−0.08 (0.22)	−0.07 (0.02)[a]	.988	.00
r_3, a_4	−0.02 (0.10)	−0.69 (0.14)[a]	−0.25 (0.21)	−0.07 (0.02)[a]	.991	.00

[a] Statistically significant at the 1 percent level.
[b] Statistically significant at the 5 percent level.
Note: The d reported is for the original equation before corrections were made for the presence of autocorrelation. Standard errors are in parentheses below coefficients.
Sources: See Table 1. Data for γ taken from U.S. Department of Commerce.

on marginal time valuation. In the recession of the 1950's, for example, the magnitudes are $6.04 billion in 1958 and $9.06 billion in 1954. The results of Table 2, therefore, suggest that the empirical magnitude of the leisure time valuation effect produced by involuntary unemployment on spending is not insignificant and indeed that such an effect complements the real-balance effect in producing a self-adjusting economic system.

IV. Conclusions and Implications

Given the assumption of normality, an increase in the wage rate always leads to an increase in income and consumption,

which in turn tends to increase transactions time. Since both leisure and consumption enter the utility function, the consumer is loath to increase transactions time at the expense of leisure. In order to maintain or even increase his leisure time the consumer is induced to hold larger money balances. This increase in money balances demanded, associated with an increase in income, is the result of a resource or income effect caused by an increase in the wage rate. This reasoning, whereby an increase in the demand for money balances is associated with an increase in transactions, yields the traditional transactions motive for holding money. In addition, there is another effect, not considered in traditional analysis, which indicates why an increase in income caused by an increase in the wage rate will lead to an increase in the demand for money. This is the leisure time valuation effect. Since a unit of leisure may be purchased by decreasing work or by increasing money balances, an increase in the wage rate results therefore in an increase in the relative price of purchasing leisure by decreasing work. Thus it follows that an increase in the wage rate will lead to an increased demand for money independent of any increase in transactions.

Considering the net result of both the resource and leisure time valuation effects, if a rise in income, wage rate constant, produced a proportionate rise in the demand for money, then the total increase in the demand for money, produced by an increase in the wage rate associated with a rise in income, should be greater than proportionate to the increase in income due to the leisure valuation effect. Thus an income elasticity (which includes both the resource and leisure valuation effects) less than unity is not necessary for the confirmation of William Baumol's and James Tobin's models which predict economies of scale in the holding of transactions balances.[9] If the leisure valuation effect is a significant determinant of the demand for money, an income elasticity of one would indicate economies of scale in transactions balances. If on the other hand, the resource effect elasticity is equal to one, then an income elasticity greater than one would be implied when both the resource and leisure valuation effects are taken into account.

When viewed in the context of disequilibrium, the leisure time valuation effect provides a theoretical explanation of the spillover effects among the money, labor, and commodity markets produced by the impact of the individual's employment condition on his valuation of time. Specifically, unemployment or overemployment affect the individual's valuation of time and affect the individual's demand for money and commodities by altering the valuation of the service money provides.

The tentative empirical results of Table 2 suggest that the relation between the unemployment rate and the demand for money is negative and statistically significant. Further, if the elasticity of the demand for money presented in Table 2 is used to estimate the magnitude of the leisure time valuation effect, the magnitudes gauged even for mild recessions appears significant. Such results suggest that the leisure time valuation effect might form a complement to the real-balance effect in producing automatic adjustment to full employment.

Given an invariant demand for money function and an exogenous money stock, a fall in prices increases real-money balances and stimulates spending. By the leisure time valuation effect, a fall in the wage

[9] Employing the standard formulation of the demand for money, which incorporates only the resource effect, Meltzer, p. 220, has concluded that an income or wealth elasticity less than unity would confirm economies of scale in holding transactions balances.

rate or a rise in unemployment will decrease the demand for money. The Invisible Hand is thus more powerful than previously supposed since changes in the wage rate and unemployment yield a new dimension to the adjustment process.

REFERENCES

R. J. Barro and H. I. Grossman, "A General Disequilibrium Model of Income and Employment," *Amer. Econ. Rev.*, Mar. 1971, *61*, 82–93.

W. J. Baumol, "The Transactions Demand for Cash: An Inventory Theoretic Approach," *Quart. J. Econ.*, Nov. 1952, *66*, 545–56.

K. Brunner and A. Meltzer, "The Uses of Money: Money in the Theory of an Exchange Economy," *Amer. Econ. Rev.*, Dec. 1971, *61*, 784–805.

G. C. Chow, "On the Long-Run and Short-Run Demand for Money," *J. Polit. Econ.*, Apr. 1966, *74*, 111–29.

D. Cochrane and G. H. Orcutt, "Application of Least Squares Regressions to Relationships Containing Auto-correlated Error Terms," *J. Amer. Statist. Assn.*, 1949, *44*, 32–61.

R. W. Goldsmith, *A Study of Saving in the United States*, IV, Princeton 1956.

———, *The National Wealth of the United States in the Postwar Period*, Princeton 1962.

A. Meltzer, "The Demand for Money: The Evidence from the Time Series," *J. Polit. Econ.*, June 1963, *71*, 219–46.

J. Niehans, "Money and Barter in General Equilibrium with Transactions Costs," *Amer. Econ. Rev.*, Dec. 1971, *61*, 773–83.

T. R. Saving, "Transactions Cost and the Demand for Money," *Amer. Econ. Rev.*, June 1971, *61*, 407–20.

J. Tobin, "The Interest Elasticity of Transactions Demand for Cash," *Rev. Econ. Statist.*, Aug. 1956, *38*, 241–47.

U.S. Bureau of Census, *Historical Statistics of the United States, Colonial Times to 1957*, Washington 1960.

U.S. Dept of Commerce, *Business Statistics, 1967*, Washington 1967.

Limit Pricing, Potential Entry, and Barriers to Entry

By David P. Baron*

A major advance in the theory of imperfect competition has been the recognition that established firms must take into account possible actions by potential as well as by existing competitors. The empirical research on potential entry has tended to focus on the effects of concentration and barriers to entry on industry profits and price levels, while theoretical study has emphasized the limit pricing behavior of established firms. The study of barriers to entry was pioneered by Joe Bain (1956), who investigated barriers created by 1) product differentiation, 2) absolute cost advantages, and 3) economies of large scale operations. Paolo Sylos-Labini emphasized the effects of returns to scale as a barrier to entry, and Franco Modigliani has provided an important interpretation of the early work of Bain and Sylos-Labini. Empirical studies by Bain and H. Michael Mann (1966) indicate a positive relationship between concentration and profitability, while Marshall Hall and Leonard Weiss found that "Concentration would seem to be less important than the capital requirements barrier as a determinant of profitability" (p. 327).

The theoretical study of the impact of potential entry was begun by Bain (1949) who first introduced the notion that potential competition may lead established firms to sacrifice current profit in order to preclude entry. The deterministic theory of potential entry and limit pricing has been advanced by Roy Harrod, Philip Andrews, H R. Edwards, J. R. Hicks, Frank Hahn, B. Peter Pashigian, Graham Pyatt, Darius Gaskins, and by Jagdish Bhagwati, who has provided a survey and analysis of much of this work.

Oliver Williamson first suggested a probabilistic approach to the issue of potential entry in which the established firms assess the probability that a potential entrant will actually enter the industry. Morton Kamien and Nancy Schwartz have developed and interpreted a stationary, probabilistic model in which an established firm chooses a limit price so as to maximize discounted expected profit over an infinite horizon with the time until entry being uncertain. In an earlier paper I analyzed the implications of alternative assumptions regarding the established firm's assessments of the probability of entry and its postentry behavior. The purpose of this paper is to investigate, using a probabilistic model, the effects of potential entry and barriers to entry on the price and profitability of established firms in an industry faced with potential entry.

I. The Model and Optimality Conditions

The model represents an established firm in an industry composed of a number of firms selling differentiated but related products.[1] Given K_n firms in the industry at the end of period $n-1$, an established firm chooses a period n limit price $p_n(K_n)$

* Associate professor of managerial economics and decision sciences, Northwestern University. Support for this study was provided by the National Science Foundation.

[1] The established firm also may be thought of as a single monopolist or as a cartel.

conditional on that number of firms. Based on the limit price and the number of firms in the industry, k_n potential entrants decide to enter the market in period n.[2] The established firm is assumed to be able to alter its price for the next period[3] and hence chooses a price $p_{n+1}(K_{n+1})$ for period $n+1$ where $K_{n+1}=K_n+k_n$. The profit of the established firm in period n is assumed to depend on the number of firms in the industry in that period with profit if no firms enter ($K_{n+1}=K_n$) denoted by $\pi_n(p_n(K_n), K_n)$. If $k_n>0$ firms enter, period n profit is assumed to be reduced. The short-term reaction of the established firm to entry may be characterized in a variety of manners such as those discussed in my earlier paper. Here, the result of that short-term reaction will be reflected in a postentry profit $\pi_n(K_{n+1}) < \pi_n(p_n(K_n), K_n)$ for all $K_{n+1}>K_n$ where $\pi_n(K_{n+1})$ does not depend on the preentry price and is assumed to be constant.[4] The postentry profit may be understood to represent the established firm's initial reaction to entry, while its complete reaction is to choose a price $p_{n+1}(K_{n+1})$ for the next period. The assumption that the profit after entry is less than that prior to entry is in the spirit of the empirical studies of Bain and Mann that indicate a positive relationship between profitability and concentration and agrees with the behavior of classical Cournot markets with entry as considered by Charles Frank, for example. In addition to the profit earned in the industry in question the established firm may earn profit from operations in other industries and that profit is denoted by R_n.

[2] Firms may enter any time before the end of the period, but to simplify the notation and the analysis all entrants will be assumed to enter at the same point in time.

[3] The length of the period may be considered to be defined by the time required for the established firm to respond to entry.

[4] Kamien and Schwartz assume that postentry profit is constant and the same in all periods after entry.

The number of firms that will decide to enter the industry in a period is uncertain, so the established firm is assumed to assess a (subjective) probability distribution on the number of potential entrants that will actually enter the market in each period. The probability of entry is a function of the assessed profitability to the potential entrant from being in the industry.[5] One possible assumption is that the probability of entry depends on the profit earned by the established firms. This assumption will not be employed, however, since it is doubtful that a potential entrant would be able to observe the profit of an established firm, particularly for a multiproduct firm. Also, the profit to an entrant may well be different from that of an established firm because of cost differences and the effect of product differentiation.[6] A potential entrant, however, is able to observe the price charged by an established firm, and that price is an important indicator of the profitability to the potential entrant. No implication is made, however, that a potential entrant believes that the current price will be in effect after entry has taken place. The potential entrant also is likely to believe that the number of firms in the industry affects profitability with a decrease in concentration reducing the likelihood that entry will be profitable.

The probability $G_n(k_n | p_n(K_n), K_n)$ that k_n firms enter in period n thus is a function of the price $p_n(K_n)$ and of the number K_n of firms in the industry. The probability of no more than a specified number m of firms entering the market is assumed to be decreasing in p_n and increasing in K_{n-1} or more formally:

[5] Roger Sherman and Thomas D. Willett utilized an expected utility maximization model for determining if a potential entrant should or should not enter.

[6] The assumption that entry depends on the preentry profit of established firms has been employed by Pyatt and Hahn. The sometimes paradoxical implications of such an assumption were explored in my earlier paper.

$$(1)\quad \sum_{k_n=0}^{m} G_n(k_n \mid p_n, K_n) < \sum_{k_n=0}^{m} G_n(k_n \mid p_n^*, K_n)$$

for $p_n > p_n^*$ and for $m=0, 1, 2, \ldots,$ $n=1, 2, \ldots$, while

$$(2)\quad \sum_{k_n=0}^{m} G_n(k_n \mid p_n, K_n) > \sum_{k_n=0}^{m} G_n(k_n \mid p_n, K_n^*)$$

for $K_n > K_n^*$ and for $m=0, 1, 2, \ldots,$ $n=1, 2, \ldots$

Consequently, as price increases entry is more likely,[7] and as the number of firms in the industry increases entry is less likely. In the terminology of stochastic dominance (see Josef Hadar and William Russell and Giora Hanoch and Haim Levy) the distribution in (1) conditional on p_n^* is dominated in the first degree by the distribution for p_n and similarly in (2) the distribution for K_n is dominated in the first degree by that for K_n^*. In order to avoid paradoxical results, it also is postulated that the industry is such that it is not optimal to encourage firms to enter in the current period in order to reduce the probability of future entry. Similarly, at the price that maximizes the current-period profit the probability of entry is taken to be positive.

The established firms are assumed to maximize the expected utility of profit over an infinite horizon with a utility function that is temporally additive. Let U_n be a monotone increasing, concave, twice continuously differentiable, cardinal utility function for period n and let $\beta \in (0, 1)$ reflect time preference. The optimal expected utility $F_n(K_n)$ for periods $n, n-1, n+2, \ldots$ depends only on the number of firms in the industry at the end of period $n-1$ and is defined as equation (3). The term $F_{n+1}(K_{n+1})$ is the optimal expected utility in periods $n+1, n+2, \cdots$, and is similarly defined. Given that profit is decreasing in the number of firms in the industry and that future entry is preferred to present entry, $F_{n+1}(K_{n+1})$ may be shown to be decreasing in the number of firms in the industry. The profit function is assumed to be concave and continuously differentiable, and for $k_n \geq 1$, G_n is assumed to be continuously differentiable. The optimal price $\hat{p}_n(K_n)$ satisfies the first-order condition[8]

[7] This probabilistic treatment of entry is equivalent to Hicks' case of oligopolistic expectations in which an increase in "close" period output was assumed to increase open period revenue by reducing the "amount" of entry. Hicks' polypolistic expectations may be represented by replacing the inequality in (1) by an equality.

[8] The second-order condition is

$$(1 - \sum G_n)(U_n''(\pi_n')^2 + U_n'\pi_n'') - 2\sum G_n' U_n' \pi_n' - \sum G_n'' M_n(K_n + k_n) < 0$$

where $M_n(K_n+k_n)$ is defined below. A sufficient condition for $\hat{p}_n(K_n)$ to yield a maximum is that $G_n' \geq 0$ and $G_n'' \geq 0$. The second-order condition is assumed to be satisfied at $\hat{p}_n(K_n)$.

$$(3)\quad F_n(K_n) \equiv \max_{P_n(K_n)} \left\{ \left(1 - \sum_{k_n=1}^{\infty} G_n(k_n \mid p_n(K_n), K_n)\right)[U_n(\pi_n(p_n(K_n), K_n) + R_n) + \beta F_{n+1}(K_n)] + \sum_{k_n=1}^{\infty} G_n(k_n \mid p_n(K_n), K_n)[U_n(\pi_n(K_n + k_n) + R_n) + \beta F_{n+1}(K_n + k_n)] \right\}$$

$$(4)\quad \left(1 - \sum_{k_n=1}^{\infty} G_n\right) U_n' \pi_n' - \sum_{k_n=1}^{\infty} G_n' [U_n(\pi_n(\hat{p}_n(K_n), K_n) + R_n) + \beta F_{n+1}(K_n) - U_n(\pi_n(K_n + k_n) + R_n) - \beta F_{n+1}(K_n + k_n)] = 0$$

where $(1 - \sum_{k_n=1}^{\infty} G_n)$ is the probability of no entry. The argument of the functions will be omitted when the meaning is clear. The term in brackets, denoted by $M_n(K_n+k_n)$ is the difference between the optimal expected utility without entry in period n and optimal expected utility with entry. Since $\pi_n(\hat{p}_n(K_n), K_n) > \pi_n(K_n+k_n)$ and $F_{n+1}(K_n) > F_n(K_n+k_n)$ for $k_n \geq 1$, $M_n(K_n+k_n)$ is positive and increasing in k_n. The change in $M_n(K_n+k_n)$ as the distribution of k_n shifts to the right (in the sense of first-degree stochastic dominance) is of the same sign as

$$\sum_{k_n=1}^{\infty} G_n' M_n(K_n + k_n) = \sum_{k_n=0}^{\infty} G_n' [U_n(\pi_n(\hat{p}_n(K_n), K_n) + R_n) + \beta F_{n+1}(K_n) - U_n(\pi_n(K_n + k_n) + R_n) - \beta F_{n+1}(K_n + k_n)]$$

where $\pi_n(K_n+k_n) \equiv \pi_n(\hat{p}_n(K_n), K_n)$ and $F_{n+1}(K_n+k_n) \equiv F_{n+1}(K_n)$ for $k_n = 0$. The shift thus implies that

$$\sum_{k_n=1}^{\infty} G_n M_n(K_n + k_n)$$

is increasing in $p_n(K_n)$ and hence that

$$\sum_{k_n=1}^{\infty} G_n' M_n(K_n + k_n) > 0$$

To satisfy (4) the optimal price is such that the probability of no entry $(1 - \sum_{k_n=1}^{\infty} G_n)$ is positive and since marginal utility is positive, marginal profit is positive. Consequently, the established firm prices below the price that maximizes preentry profit[9] and prices less than the price that makes the probability of entry equal to one.[10] These results hold for any monotone increasing utility function and thus are due to the probabilistic nature of entry and not to the risk preferences reflected by the utility function.

Harrod suggested that if entry is easy the established firm will price equal to average cost in order to prevent entry. If the firm follows this practice in every period, there is no difference between the utility with and without entry, so $M_n(K_n) = 0$. The optimal limit price thus is higher than the price that equals average cost unless, of course, marginal profit is zero at that price. The probabilistic equivalent of Harrod's entry preventing price may be defined as the price $p_n^e(K_n)$ that makes the probability of no entry equal to one or $G_n(k_n=0 \mid p_n^e(K_n), K_n) \equiv 1$. The optimal limit price is at least as great as the entry-preventing price, since otherwise an increase in price would increase profit and leave the probability of entry equal to zero. This result is a prediction of Dale Osborne's finding that entry has occurred in industries using limit pricing. The deterministic entry-preventing price may also be thought of as the lowest price the established firm could set without finding it preferrable to leave the industry. One such price $p_n^*(K_n)$ satisfies

$$(5)\quad (1 - \sum G_n) U_n(\pi_n(p_n^*(K_n), K_n) + R_n) + \sum G_n U_n(\pi_n(K_n + k_n) + R_n) \equiv U_n(R_n)$$

[9] Kamien and Schwartz obtained the same result in a stationary, probabilistic model. They also demonstrate that the preentry profit is constant over time, and a similar result obtains for the model here if U_n, G_n, and R_n are the same in every period.

[10] Hicks' polypolistic expectations implies that $G_n' = 0$ for all k_n, and then the established firm maximizes preentry profit.

The entry-preventing price may be less than $p_n^*(K_n)$, so the established firm may find it impossible to stay in the industry and prevent entry.

II. Risk Aversion and the Optimal Limit Price

The established firm's limit pricing decision involves the risk that profit will be reduced by entry, and the risk preferences reflected in the firm's utility functions U_n affect the level of the limit price. Risk preferences will be measured by the Arrow-Pratt index of absolute risk aversion $r_{U_n}(y) = -U_n''(y)/U_n'(y)$, where U_n'' and U_n' are the second and first derivatives, respectively. The interpretation of the index is that if $r_{U_n}(y)$ increases for all y, as with a shift in risk preferences, the certainty equivalent associated with the risk increases. The certainty equivalent CE_n of period n profit is defined by

$$(6)\quad U_n(CE_n+R_n) \equiv \left(1-\sum_{k_n=1}^{\infty} G_n\right)U_n(\pi_n(\hat{p}_n(K_n), K_n) + R_n) + \sum_{k_n=1}^{\infty} G_n U_n(\pi_n(K_n+k_n)+R_n)$$

and is the minimum amount the firm would accept in exchange for the period n *ex ante* profit.

Given certain properties of the probability distribution, the optimal period n price may be shown to be a decreasing function of risk aversion. That is, let U_n^1 and U_n^2 be two utility functions such that $r_{U_n^1}(y) \geq r_{U_n^2}(y)$ for all y and $r_{U_n^1} > r_{U_n^2}(y)$ for some y with positive probability, and let $\hat{p}_n^1(K_n)>0$ be optimal for U_n^1 and $\hat{p}_n^2(K_n)>0$ be optimal for U_n^2. Rewrite (4) for U_n^2 evaluated at $\hat{p}_n^2(K_n)$ as

$$(7)\quad (1-\sum G_n)\pi_n' - \sum G_n' \phi_n^2(k_n) = 0$$

where $\phi_n^2(k_n) \equiv M_n^2(K_n+k_n)/U_n^{2\prime}(y^2)$ and $y^2 = \pi_n(\hat{p}_n^2(K_n), K_n)+R_n$. John W. Pratt (equation 22, p. 129) has shown that if $r_{U_n^1}(y) > r_{U_n^2}(y)$ for all y, then $(U_n^1(w) - U_n^1(v))/U_n^{1\prime}(y) > (U_n^2(w)-U_n^2(v))/U_n^{2\prime}(y)$ for $y \geq w > v$. To maintain the intertemporal relationship between U_n and U_j, $j\neq n$, scale the utility function U_n^1 so that $U_n^{2\prime}(y^2) = U_n^{1\prime}(y^2)$. An increase in risk aversion thus implies that $\phi_n^1(k_n) > \phi_n^2(k_n)$ for $k_n \geq 1$. A sufficient, but not necessary, condition for risk aversion to decrease the optimal limit price is that $G_n' \geq 0$ for all $k_n \geq 1$,[11] since then (7) for U_n^1 is negative evaluated at $\hat{p}_n^2(K_n)$, so $\hat{p}_n^2(K_n) > \hat{p}_n^1(K_n)$. A necessary condition is that $\Sigma G_n'[\phi_n^1(k_n) - \phi_n^2(k_n)] \,|\, \hat{p}_n^2(k_n) > 0$. To simplify the following discussion, G_n' is assumed to be nonnegative which implies that an increase in prices increases the probability that k_n firms enter, $k_n \geq 1$.

The interpretation of this result is that as the established firm becomes more risk averse it prefers to exchange period n profit for a lower probability of entry and thus reduces price in period n. The risk to the established firm is that potential entrants will actually enter, and the more averse to risk the firm becomes, the more willing it is to sacrifice current profit for future profit. The effect of risk aversion on the limit price makes intuitive sense and may help explain the differences in profit between industries found by Bain, Mann, and Irving Fisher and George Hall. Certainly many other factors such as those considered in the following section on barriers to entry are important as determinants of the limit price, but the analysis here indicates that risk aversion may act to lower the limit price and thus has an effect similar to a barrier to entry in that both tend to make entry less likely. Increased risk aversion also reduces the certainty equivalent.

The effect of profits from other ac-

[11] The condition that $G_n' \geq 0$ for $k_n \geq 1$ implies, but is not implied by, first-degree stochastic dominance.

tivities R_N on the limit price may be determined directly from the risk aversion result. A utility function is said to exhibit decreasing (increasing) (constant) absolute risk aversion if $r_{U_n}(y)$ is a decreasing (increasing) (constant) function of y. An increase in R_n for U_n exhibiting decreasing (increasing) (constant) absolute risk aversion consequently results in an increase (decrease) (no change) in the optimal limit price.[12] Kenneth Arrow has argued that decreasing absolute risk aversion is a reasonable behavioral assumption. For U_n decreasingly absolute risk averse an increase in R_n decreases risk aversion, and the firm is willing to accept a higher probability of entry in exchange for greater profit if no entry occurs.[13]

III. Barriers to Entry and Concentration

The principal barriers to entry considered by Bain are 1) economies of large scale operations, 2) absolute cost advantages, and 3) product differentiation advantages. One general way to consider the effects of barriers to entry on the limit price is to assume that they are reflected by the established firm's assessment of the probability of entry. An increase in the height of barriers to entry will be represented by a shift to the left in the distribution function of k_n for all prices and for all K_n. For a given $p_n(K_n)$ and K_n, higher barriers to entry thus are assumed to increase the probability that fewer than m firms enter for $m=1, 2, \ldots$, and hence result in a distribution dominated in the first degree by the distribution for lower barriers to entry. Such a change in the probability distribution of k_n affects both the established firm's certainty equivalent and its optimal price. Since period n utility and future expected utility $F_{n+1}(K_{n+1})$ are decreasing in k_n, an increase in the height of the barriers to entry increases the expected utility in (3). The established firm would then be less likely to leave the industry and would require a greater payment for the sale of its activities in the industry. Empirical work by Bain and Mann supports the hypothesis that average profits are higher with "very high barriers" than with lower barriers. In the context of the probabilistic model considered here expected utility and the certainty equivalent profit increase as barriers become "higher."

While higher barriers to entry increase the expected utility of the established firm, the optimal price may increase or decrease. In (4) the probability of no entry $(1-\Sigma G_n)$ increases with higher barriers to entry but the G_n' terms also are affected. If the $G_n'/(1-\Sigma G_n)$, $k_n \geq 1$, terms are not increased by the higher barriers to entry, then (4) evaluated at the price optimal with lower barriers to entry is positive and the optimal price is increased. The term $G_n'/(1-\Sigma G_n)$ is referred to as the "hazard rate" in reliability theory, and if an increase in the barriers to entry reduces the hazard (of entry) rate, the optimal price increases.[14] Bain (1956) has argued that the higher are the barriers to entry the closer the limit price is to the profit maximizing price, and for the model considered here a reduction in the hazard rate for all $k_n \geq 1$ is a sufficient, but not a necessary, condition for this result.

Concentration in the industry affects the established firm by influencing profit and the probability of entry. Holding profit constant, an increase in concentra-

[12] This result obtains by letting $U_n^1(y)=U_n(y)$ and $U_n^2(y)=U_n(y+\Delta y)$, $\Delta y>0$, and using the preceding analysis.

[13] If, for example, a potential entrant must pay to an established firm an entry fee in the form of the purchase of the rights to a patented process, R_n is increased and the firm with decreasingly absolute risk averse risk preferences will raise its limit price which increases the probability of further entry, *ceteris paribus*.

[14] Kamien and Schwartz assume that the hazard rate is increasing in the price charged by the established firm.

tion, such as would occur from firms leaving the industry, affects the probability of entry by shifting the distribution function of the number of entrants to the right as indicated in (2). Such a shift decreases expected utility in agreement with the studies of Bain, Mann, and S. A. Rhoades and decreases the limit price if the hazard rate is increased. Changes in concentration also affect profit, however, so the net effect of exit from the industry is difficult to determine without further assumptions.

The effect of economies of large scale operations on entry may also be reflected by changes in the probability of entry. Modigliani concluded that "[the 'highest entry-preventing price'] will tend to be higher the steeper the cost curve, that is the greater the economies of scale" (p. 220). Similarly, Hall and Weiss found that profitability is positively related to the capital requirements barrier to entry. The effect of such a barrier is to shift to the left the distribution function of the number of firms entering. The shift increases the expected utility giving a theoretical prediction of Hall and Weiss' findings. If the shift reduces the hazard rate, the limit price increases as Modigliani suggested.

Absolute cost advantages may have two types of effects on the pricing policies of an established firm. First, the knowledge on the part of potential entrants that an established firm has achieved a cost reduction, through a technological advance, for example, may reduce the assessed probability of entry. Second, a reduction in cost to an established firm may affect the limit price independent of the probability distribution. For example, suppose that an established firm achieves a reduction in fixed costs in period n but that the potential entrants are unaware of it. A reduction in fixed costs in period n has the same effect as an increase in profits R_n from other activities, so a reduction in fixed costs results in an increase (decrease) (no change) in the optimal limit price in period n if the utility function U_n exhibits decreasing (increasing) (constant) absolute risk aversion. A decrease in a constant marginal cost has both a risk aversion effect and a "cost effect." This risk aversion effect occurs because for a fixed output a reduction in marginal cost is equivalent to a decrease in fixed cost. As in the deterministic theory of the firm, the cost effect tends to decrease price by altering the relationship between marginal revenue (utility) and marginal cost. For nondecreasing absolute risk aversion, a decrease in a constant marginal cost results in a decrease in the limit price,[15] but with decreasing absolute risk aversion the limit price may increase if the risk aversion effect exceeds the cost effect.

Bain found that product differentiation, primarily in the form of advertising, played the most important role in building and sustaining barriers to entry in consumer industries. Utilizing assumptions similar to those for price,[16] the optimal product differentiation expenditures may be shown to be greater than the expenditures that maximize current period profit. The firm thus again is willing to exchange current profit for a lower probability of entry. If product differentiation expenditures decrease the hazard rate in (4), the optimal limit price increases.

The impact of industry growth on the behavior of both established firms and potential entrants is an important issue in industrial organization, and in the context of the model presented in this paper growth may affect both profit and the assessed probability of entry. Since shifts in the distribution function of k_n have al-

[15] This result requires that postentry output is less than preentry output.

[16] An increase in product differentiation expenditures is assumed to shift the probability distribution of the number of firms that enter to the left.

ready been considered, only changes in future profit will be considered at this point with the probability distribution for a given price and K_n assumed fixed. If growth increases the difference in future expected utility with and without entry ($[F_{n+1}(K_n)-F_{n+1}(K_n+k_n)]$) and $G'_n \geq 0$ for $k_n \geq 1$, the term $\Sigma G'_n[F_{n+1}(K_n) - F_{n+1}(K_n+k_n)]$ increases and the optimal limit price in period n is decreased to reduce the probability of future entry. If growth lessens the difference, the optimal price will be increased, since entry causes less of a reduction in profitability.

IV. Conclusions

Potential entry has been represented in a probabilistic manner with the probability of entry dependent on the price charged by an established firm and on the number of firms in the industry. With this probabilistic view of entry the optimal limit price is less than the price that maximizes preentry profit and is greater than the price that equates price and average cost. Since the limit pricing decision involves uncertain profits, it seems natural to ask how risk preferences affect the limit price. Given certain properties of the probability distribution, an increase in risk aversion in period n as measured by the Arrow-Pratt index results in a reduction in the period n limit price, since the firm is willing to accept a reduced current period profit in exchange for a lower probability of entry. Barriers to entry may be reflected in the assessed probability of entry with higher barriers being represented by a shift to the left in the distribution function of the number of actual entrants. Higher barriers to entry result in an increased expected utility, and if the hazard rate is reduced, the optimal limit price is increased.

The empirical research on barriers to entry, concentration, and risk adjusted profitability has been characterized by debates regarding what it is that is actually being measured. For example, see the recent papers by Fisher and Hall, Yale Brozen, Mann (1969), and Richard Caves and B. S. Yamey. The results contained herein indicate the individual effects of concentration, barriers to entry, probability assessments, and risk aversion on profitability and price, but attempting to identify and isolate individual effects from empirical data is likely to be extremely difficult.

REFERENCES

P. W. S. Andrews, "A Reconsideration of the Theory of Individual Business," *Oxford Econ. Pap.*, Jan. 1949, *1*, 54–89.

K. J. Arrow, *Essays in the Theory of Risk-Bearing*, Markham 1971.

J. S. Bain, "A Note on Pricing in Monopoly and Oligopoly," *Amer. Econ. Rev.*, Mar. 1949, *39*, 448–64.

———, *Barriers to New Competition*, Cambridge, Mass. 1956.

D. P. Baron, "Limit Pricing and Models of Potential Entry," *Western Econ. J.*, Sept. 1972, *10*, 298–307.

J. D. Bhagwati, "Oligopoly Theory, Entry-Prevention, and Growth," *Oxford Econ. Pap.*, Nov. 1970, 297–310.

Y. Brozen, (1969a) "Significance of Profit Data for Antitrust Policy," *Antitrust Bull.*, spring 1969, *14*, 119–39.

———, (1969b) "Barriers Facilitate Entry," *Antitrust Bull.*, winter 1969, *14*, 851–54.

R. E. Caves and B. S. Yamey, "Risk and Corporate Rates of Return: Comment," *Quart. J. Econ.*, Aug. 1971, *85*, 513–17.

H. R. Edwards, "Price Formation in Manufacturing Industry and Excess Capacity," *Oxford Econ. Pap.*, Feb. 1955, *7*, 94–118.

I. N. Fisher and G. R. Hall, "Risk and Corporate Rates of Return," *Quart. J. Econ.*, Feb. 1969, *83*, 79–92.

——— and ———, "Risk and Corporate Rates of Return: Reply," *Quart. J. Econ.*, Aug. 1971, *85*, 518–22.

C. R. Frank, Jr., "Entry in a Cournot Market," *Rev. Econ. Stud.*, July 1965, *32*, 245–50.

D. W. Gaskins, Jr., "Dynamic Limit Pricing: Optimal Pricing Under Threat of Entry," *J. Econ. Theory*, Sept. 1971, *3*, 306–22.

J. Hadar and W. R. Russell, "Rules for Ordering Uncertain Prospects," *Amer. Econ. Rev.*, Mar. 1969, *59*, 25–34.

F. H. Hahn, "Excess Capacity and Imperfect Competition," *Oxford Econ. Pap.*, Oct. 1955, *7*, 229–40.

M. Hall and L. Weiss, "Firm Size and Profitability," *Rev. Econ. Statist.*, Aug. 1967, *49*, 319–30.

G. Hanoch and H. Levy, "The Efficiency Analysis of Choices Involving Risk," *Rev. Econ. Stud.*, July 1969, *36*, 335–46.

R. F. Harrod, *Economic Essays*, New York 1952, 139–87.

J. R. Hicks, "The Process of Imperfect Competition," *Oxford Econ. Pap.*, Feb. 1954, *6*, 41–54.

M. I. Kamien and N. L. Schwartz, "Limit Pricing and Uncertain Entry," *Econometrica*, 1971, *39*, 441–54.

H. M. Mann, "Seller Concentration, Barriers to Entry, and Rates of Return in Thirty Industries, 1959–1960," *Rev. Econ. Statist.* Aug. 1966, 296–307.

———, "A Note on Barriers to Entry and Long Run Profitability," *Antitrust Bull.*, winter 1969, *14*, 845–49.

F. Modigliani, "New Developments on the Oligopoly Front," *J. Polit. Econ.*, June 1958, *66*, 215–32.

D. K. Osborne, "The Role of Entry in Oligopoly Theory," *J. Polit. Econ.*, Aug. 1964, *72*, 396–402.

B. P. Pashigian, "Limit Price and the Market Share of the Leading Firm," *J. Ind. Econ.*, July 1968, *16*, 165–77.

J. W. Pratt, "Risk Aversion in the Small and in the Large," *Econometrica*, Jan./Apr. 1964, *32*, 122–36.

G. Pyatt, ' Profit Maximization and the Threat of New Entry," *Econ. J.*, June 1971, *81*, 242–55.

S. A. Rhoades, "Concentration, Barriers and Rates of Return: A Note," *J. Ind. Econ.*, Nov. 1970, *19*, 82–88.

R. Sherman and T. D. Willett, "Potential Entrants Discourage Entry," *J. Polit. Econ.*, Aug. 1967, *75*, 400–03.

P. Sylos-Labini, *Oligopoly and Technical Progress*, Cambridge, Mass. 1962.

O. E. Williamson, "Selling Costs as a Barrier to Entry," *Quart. J. Econ.*, Feb. 1963, *77*, 112–28.

Wage Determination, Inflation, and the Industrial Structure

By STEPHEN A. ROSS AND MICHAEL L. WACHTER*

The standard analysis of inflation is based on the existence of a short- and long-run Phillips curve or tradeoff between the rate of inflation and unemployment. This tradeoff results from the effects of excess demand pressure on a labor market characterized by money illusion on the part of workers and firms, or by institutional rigidities imposed by labor unions and producer monopolies.[1] Demand-pull and cost-push elements interact in a complex disequilibrium process, and little attempt has been made to derive rigorously the dynamics or the properties of the steady state. Neoclassical economists have attacked this formulation of the Phillips curve largely on the basis of their own analysis of the long-run equilibrium properties of the system.[2] They agree that the Phillips curve may exist in the short run but insist that it does not exist (that is, it is vertical) in the long run.[3] In the long-run expectations are realized so that money illusion is unsustainable, and if the degree and strength of noncompetitive elements are unchanged, they argue that there can be no tradeoff between inflation and unemployment.

It is also argued that monopoly power cannot result in inflation. In particular, unions desire a relative wage advantage over nonunion workers, and monopolies desire a certain profit markup. In equilibrium, both unions and monopolies will have exercised the full power of their noncompetitive advantage. The wage and profit markups are then fixed in real terms, and there will be no incentive to further increase relative wages and prices. As a result the rate of inflation will not be influenced by an *unchanging* level of noncompetitive forces.[4] As stated by Phelps:

> An increase of monopoly power—due, say, to increased concentration—will raise prices relative to wages at any given unemployment rate and productivity level; but once, at the prevailing unemployment rate, the real wage has fallen (relative to productivity) enough

* Associate professors of economics, University of Pennsylvania. The research was supported by a grant from the National Science Foundation.

[1] The present form of the Phillips curve is due largely to the work of George Perry. More recent research, which attempts to refute the neoclassical model and to utilize equations similar to those of Perry, includes the papers by Robert Solow and Robert Gordon.

[2] The neoclassical position is described in the collection of papers in Edmund S. Phelps. See also the article by Milton Friedman.

[3] Attempts to statistically verify the existence of a vertical long-run Phillips curve have largely been unsuccessful, but the new theory continues to gain in popularity. The reasons for this are obvious. First, and most importantly, the standard Phillips curve analysis with its bias towards cost-push explanations of inflation lacks a strong theoretical bias. Secondly, statistical findings are never unequivocal, and since current time-series data do not include any periods of extended inflation, statistical analysis would be biased towards the observation of a long-run negatively sloped Phillips curve. In addition, no specific theory exists to suggest the exact form of the wage equation, so it is not possible to construct an agreed upon test of the neoclassical model.

[4] The neoclassical position is often taken a step further. Although it is agreed that money illusion is a cause of a short-run Phillips curve, the importance of monopoly elements (either on the part of labor or capital) that are of fixed strength is deemphasized even in the short-run tradeoff. Instead, the observation of seemingly noncompetitive behavior is attributed to the cost of information or transactions costs. These costs, like money illusion, are only a short-run phenomenon so that the long-run tradeoff between inflation and unemployment is still vertical.

> to accommodate the higher markup; this process will stop and any continuation of inflation will depend on other sources. [p. 126]

The purpose of this paper is to outline a new view of the relation between inflation and unemployment in a high employment economy. A basic innovation of our approach is that it incorporates noncompetitive behavior and some of the most important institutional features of the modern economy in a systematic manner into the statics and dynamics of the analysis. The presence of noncompetitive elements in a sector of the economy will be shown to provide a new justification for the existence of a short- and long-run tradeoff between inflation and unemployment. The long-run relationship, however, is positive when there is inflation but negative with deflation, which is in sharp contrast to both the neoclassical and the traditional Phillips curve approaches to the subject.

The analysis also gives rise to a number of additional points of interest from both a theoretical and policy viewpoint. First, the stability of the labor market and the aggregate wage-price structure is altered in a complex manner. Although the long-run Phillips curve is c-shaped, to the extent that policy makers understand the workings of the system, the short-run, negatively sloped Phillips curve actually has considerably more stability than it would if based solely on money illusion and information costs. The basis of this increased stability may be important in explaining the inability of empirical workers to find anything but a negatively sloping Phillips curve. On the other hand, the possibility of policy error is magnified, and this leads to instability. Second, the existence of a sector that pays a wage premium increases the amount of voluntary unemployment in the system and causes a redistribution of income from those not employed to those employed in the high wage sector. This type of unemployment is theoretically similar to the unemployment which arises in "dual labor" market theories. The market mechanism embodied in the dual labor market is currently popular in describing developed as well as underdeveloped economies.[5] Third, noncompetitive factors are shown to have an important dynamic influence upon inflation.

The outline of the paper is as follows. Section I develops a theory of the administered, or noncompetitive, firm. Section II introduces the competitive sector. Section III aggregates the results of the previous two sections into a single macro-economic model, and Section IV analyzes the short- and long-run dynamic behavior of this model. Section V, a summary, discusses the theoretical and policy implications of the model.

I. The Administered Sector

Nature of the Administered Firm

A maintained assumption of this paper is that firms differ in the discretion they can exercise in setting price. These differences are due to factors such as the firm's technology (as it influences the shape of its cost curves), the size of the market, and, in general, all factors that help to determine the number of firms in the industry and the degree of constraint on further entry. For expositional simplicity we divide the economy into two polar cases, those firms with pricing discretion (the noncompetitive or administered sector) and those firms that are purely price takers (the competitive sector).

In the industries with price discretion

[5] For a discussion of dual labor market theories in underdeveloped economies see W. Arthur Lewis, Michael Todaro, and John Harris and Todaro. Similar types of models for developed economies have been investigated by Peter Doeringer and Michael Piore and Bennett Harrison.

it is likely that there is some price structure, different from the price structure that would prevail if the sector were competitive, that maximizes each firm's profits.[6] Once agreement within an industry is reached on a price structure, it is then clearly both costly, in terms of agreeing on a new price structure, and risky to alter the structure. The primary risk, of course, is that of a loss of monopoly profits from a breakdown in the informal price agreements. It follows that from the industry's point of view, it is undesirable to have prices change with the frequent short-run changes in demand or cost conditions since such changes would open the possibility of disagreement and threaten the structure in an unpredictable fashion.[7]

As a consequence, a crucial component of noncompetitive, or what we will hereafter call administered pricing, is that it exhibits a certain amount of price stability in the face of variations in market conditions. To represent this we assume that the price decision is made only at the beginning of a finite planning period. During the planning period prices then either remain constant or change according to a set strategy that is part of the initial price decision. (For simplicity, our analysis concentrates on that case where prices are constant over the planning period.)

In order to preserve the industry price structure, it is necessary for the firms to develop an agreed upon method of signaling when prices should change in response to changing demand and cost conditions. Since price changes are costly and, if erratic, carry the risk of substantial loss, such a device must signal both infrequently and reliably. A traditional method for achieving this reliability is to base prices on a markup on certain standard costs that are both important in total industry variable costs and nearly identical for all firms in the industry. Thus, prices are often based on a markup on wages. Unions can be helpful if industry-wide settlements are the rule although the mere existence of a long-term wage contract is not sufficient to ensure the desired wage stability.[8] For a long-term wage contract to have operational significance, it is necessary that the contract include a wage premium that will insulate the firm from short-run changes in the economy-wide demand for labor.[9] If this were not

[6] A major reference on the existence and behavior of a noncompetitive sector is by Frederick Scherer. The notion that each firm maximizes profits is, of course, ambiguous in the oligopoly context. There are many approaches to the oligopoly problem. Most, however, are compatible with our analysis. It should also be noted that a price structure can refer to both interfirm prices and interproduct prices. In the first case, firms may not all have the same prices; however, differentials are agreed upon. The more traditional price structure refers to the price list for the various products that the firms produce.

[7] The stability of administered firm prices in the face of fluctuations in product demand is discussed by Otto Eckstein and Gary Fromm. The debate over the question of whether or not administered prices are stable dates back to the 1930's. Scherer concludes that the evidence does tend to support those who suggest that administered prices are stable. More recently, George Stigler and James Kindahl have questioned this conclusion. They contend that the empirical support for stable administered prices is based on wholesale price indices, which are seller list prices, and argue that a series created by observing the actual buying price would show a great deal more cyclical variation. This is largely due to price shading during recessions. The Stigler-Kindahl argument does not necessarily weaken the case of the administered sector presented in this paper. First, Stigler and Kindahl show only that administered prices are not stable; they do not answer the more important question (for our model) as to whether those prices are as relatively unstable as competitive prices. Secondly, it is likely that administered industries do not simply establish a single list price. Rather, they establish a trading range which has the list price as a marker. Some price shading in this framework would be within the bounds of the price agreement of the administered firms. This could account for the Stigler-Kindahl findings and would not be in contradiction to the establishment of a fixed pricing *strategy*.

[8] The term industry-wide settlements includes the case where a single contract for an industry is negotiated and the case where spillover effects link one contract in an industry to another.

[9] For an earlier and somewhat different view of the

the case, wage rates would have to be adjusted frequently to reflect market forces, and the advantage of stable administered prices based on a markup on fixed standard costs would be lost. On the other hand, with a wage premium the administered firm hires workers off a queue. Fluctuations in product demand, and thus in the derived demand for labor, cause flunctuations in numbers hired off the queue but not in the size of the money wage rate offered.[10]

The Wage and Price Decision

For simplicity and in order to concentrate on the inflation problem we assume that all variable costs are labor costs in both the administered and the competitive sectors and that over the relevant production range changes in output are simply proportional to employment. One consequence of this assumption is that since the competitive real wage, ω, will equal labor's marginal product (in the competitive sector), ω will be constant.[11] Since wages and prices change together, we also assume that the firm simultaneously makes both a wage and a price decision at the beginning of a fixed planning period, at the time T, and maintains both until the start of the next period, at time T+1.[12]

The administered firm operates under a supply and demand (for labor) constraint that, for any given moment within the planning period, τ, may be written

$$\text{(1a)} \quad L \leq L^s\left(\frac{w_T^a}{w_{T+\tau}^c}\right), \qquad 0 \leq \tau < 1$$

and

$$\text{(1b)} \quad L \leq L^d\left(\frac{p_T^a}{p_{T+\tau}^c}\right), \qquad 0 \leq \tau < 1$$

where L is the actual employment of the administered firm, w_T^a and p_T^a are its respective wage and price decisions made at time T and lasting until T+1, and $w_{T+\tau}^c$ and $p_{T+\tau}^c$ are the competitive wage and price levels at time T+τ, respectively.[13] The administered firm regards w^c and p^c as given. The first inequality denotes the labor supply limitation on hiring, and the second constraint asserts that the firm would not hire labor in excess of the amount required to produce the current product demand L^d.[14]

impact of union contracts on wage determination see the article by Eckstein and Thomas Wilson. The interaction between noncompetitive elements in product and labor markets is developed by John Dunlop (1944), (1957).

[10] The argument may be restated to emphasize that administered firms desire an assured supply of labor. This is similar to the model of administered firm behavior suggested by John Kenneth Galbraith. There, an assured labor supply as well as an assured supply of all inputs (and product demand to the extent that it is possible) is desired in order to reduce uncertainty. Galbraith interprets this as based on nonprofit maximization. We suggest that profit maximizing behavior is still relevant, but must be altered by the explicit recognition of uncertainty.

[11] The assumption of an exogenous real wage is made not only for simplicity, but also to stress the distinction between our findings and Patinkin-like disequilibrium analysis. The movement of real wages over the cycle is of great interest, however, its solution involves a more complex development of the disequilibrium behavior of the competitive sector. See, for example, Don Patinkin, R. Clower, Robert Barro and Herschel Grossman and E. Roy Weintraub and Sidney Weintraub. The marriage of our work with these competitive disequilibrium models is a difficult and distinct project.

[12] In a recent paper, George Akerlof develops a model of an economy with two noncompetitive firms where firm 1 sets its wage and price at time t, t+2, t+4, . . . , while firm 2 sets its wage and price at t+1, t+3, t+5, ... Although there are important differences, his model is the closest to our own of any in the literature.

[13] These can be interpreted either as stock or flow constraints. Since our analysis will hold in either case, we will not be explicit.

[14] This is, of course, an assumption rather than a true constraint and serves primarily to abstract from inventory accumulation. The assumption is made solely to facilitate the analysis. A stock of inventories has a role similar to that of the labor queue. Each is useful in avoiding a customer queue, and an administered firm is likely to maintain both inventories and a labor queue. An unexpected increase in the product demand would, therefore, lead to a reduction in inventories and queue size.

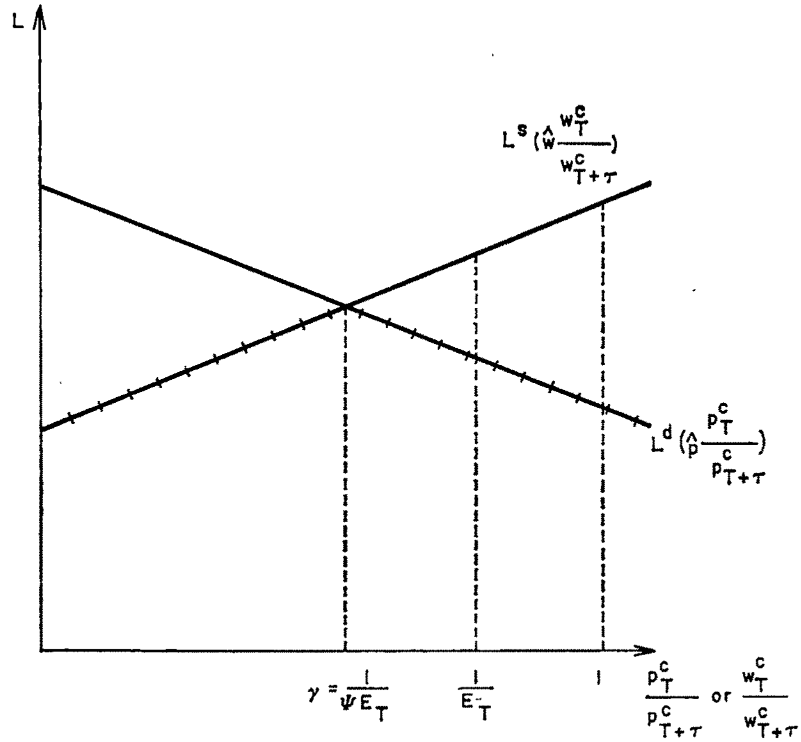


FIGURE 1

The labor demand function of the administered firm varies inversely with its relative price p^a/p^c, where p^c is used as a proxy variable for the overall price level.[15] The existence of a positive cross-price elasticity between the representative administered firm and other firms in the administered sector and/or firms in the competitive sector can justify the downward slope. The labor supply function depends positively on the administered firm's relative wage, w^a/w^c, because of the short-run immobility of labor.

Denoting the relative wage and price by $\hat{w} \equiv w^a_T/w^c_T$ and $\hat{p} \equiv p^a_T/p^c_T$, equations (1a) and (1b) may be rewritten as

$$L \leq L^s\left(\hat{w} \cdot \frac{w^c_T}{w^c_{T+\tau}}\right) \tag{1a'}$$

and

$$L \leq L^d\left(\hat{p} \cdot \frac{p^c_T}{p^c_{T+\tau}}\right) \tag{1b'}$$

The supply and demand functions in constraints (1a′) and (1b′) are depicted in Figure 1. The horizontal axis contains either $p^c_T/p^c_{T+\tau}$ or $w^c_T/w^c_{T+\tau}$; the two are equal by the constancy of the real wage. The area below the thatched line displays the constraints of (1a) and (1b).

At the beginning of each planning period the firm sets w^a_T and p^a_T. In so doing it fixes $\hat{w}$ and $\hat{p}$, the slopes of the demand and supply curves in Figure 1, and thus their intersection point γ. To the left of γ, customers queue, and to the right of γ, workers queue. At the beginning of the

[15] It might seem better to use an average administered sector price or an industry price as the denominator of the relative price term, but as will be seen later, the use of the competitive price does provide a satisfactory proxy for the relative price term and greatly simplifies the model.

planning period, τ equals zero and the administered firm is at the unity point on the horizontal axis. As τ varies from zero to unity, $p^c_T/p^c_{T+\tau}$ and $w^c_T/w^c_{T+\tau}$ may change, thus altering the position on the horizontal axis. The firm's optimal choice of γ is determined by the firm's expected rate of inflation and its presumed evaluation of profits when either workers or customers are forced to queue. If the expected rate of inflation, $E_T \equiv$ expected (p^c_{T+1}/p^c_T), were certain the firm would set $\hat{w}$ and $\hat{p}$ (the slopes of the supply and demand curves, respectively), so that $\gamma = 1/E_T$. When the forecast, E_T, is not certain, the relative cost to the firm of worker and customer queues is important in determining γ. If the realized rate of inflation is less than E_T, then workers queue, and the profit loss to the firm consists of the excess wage premium. On the other hand, when the realized inflation goes beyond E_T, the profit loss is in excess of the current profit opportunity loss of the low price. The reason for the extra loss is that disappointed customers who are forced to queue may not remain as customers in the future. The reduction in current accounting profits, therefore, understates the loss in discounted lifetime profit. Hence, the opportunity loss from underestimating the rate of inflation exceeds that of overestimating, and the risk-averse firm will pay a wage premium. This safety factor is measured by the term $\Psi > 1$ and the firm sets $\gamma = 1/E_T\Psi$.

To summarize, even though the firm expects an inflation rate of E_T, risk aversion will lead it to set its market clearing point γ so as to have a worker queue if expectations are realized. It does this to reduce the probability of having a customer queue which it views as being more costly than a worker queue.[16]

It can be shown formally that the above model of the administered firm leads to wage and price equations of the form:[17]

$$w^a_T = \phi\Psi E_T w^c_T \tag{2a}$$

$$p^a_T = \mu\Psi E_T p^c_T \tag{2b}$$

The terms $\phi\Psi$ and $\mu\Psi$ represent the markup of the administered firm since, if they were equal to unity, the administered wage and price would be equal to the competitive wage and price adjusted for the expected rate of inflation over the planning period. The basic noncompetitive features of the product and labor markets of the administered firms are represented by μ and ϕ. The variables μ and ϕ are (increasing) functions of aggregate demand and may be greater or less than unity depending upon the exact shapes of the supply and demand functions the firm faces.[18]

[16] An additional reason for the administered firm to pay a wage premium stems from the observation that the supply curve of labor is likely to be more elastic than the demand curve of labor. The firm belongs to an oligopolistic industry and is a participant to an administered pricing scheme where relative prices among firms in an industry are largely constant. These features decrease the elasticity of the product demand and thus the derived demand for labor. On the other hand, the firm may be in a perfectly competive labor market. In this case as soon as $w^c_{T+\tau} > w^a_T$ its labor supply leaves for the competitive sector, and the profits of the firm fall to zero. To avoid this the administered firm will establish a worker queue by setting $\Psi > 1$. More generally, labor supply may be sufficiently more elastic than labor demand and the argument will still be valid.

[17] The formulas are derived in our 1972 paper. Markup formulas also play a central role in the macromodel developed by S. Weintraub.

[18] If ϕ is small enough, $\phi\Psi$ may be below unity so that the noncompetitive firm would not pay a wage premium over the competitive wage. (If one assumes that the competitive and noncompetitive firms produce different products, then the notion of a price premium is not well defined and, therefore, does not play an important role in the model. Since we are assuming homogeneous labor, however, the notion of a wage premium is well defined.) The question as to whether noncompetitive firms pay a wage above the competitive wage, however, is largely empirical and should not be confused with the fact that the administered firm always pays a premium over the wage necessary to clear its own labor market. Since there is considerable evidence that noncompetitive firms do pay higher wages than competitive firms, we will assume that this is the case throughout the re

In the above discussion, the justification for the behavioral assumptions of the administered sector is based on decision variables under the control of the administered firm. The model, however, should be interpreted as being relevant to a broader group of industries than those traditionally recognized as having very high concentration ratios. First, similar results may be obtained for industries where the decision variables are, at least partially, controlled by labor unions. The results above suggest that the firms set their wage and price strategies at the beginning of a planning period (maintain a fixed strategy over that period) and pay a wage premium. Clearly, these same features are desired by labor unions. Traditionally, unions are viewed as attempting to achieve a wage for their members which is greater than the wage dictated by market conditions. Furthermore, since contract setting is a costly process, unions do not seek continuous renegotiation of contracts; rather, they prefer contracts which are fixed for a period of time. Secondly, the dichotomy of firms as competitive or noncompetitive is not critical. More realistically, industries may be classified along a continuum with respect to their noncompetitive features. The interaction among firms along the continuum will lead to the same qualitative macro-economic behavior that is described below for the dichotomous case.

II. The Competitive Sector

In the competitive sector, the wage and price adjust continuously to clear the market.[19] The assumption of instantaneous reaction in the competitive sector is made solely to concentrate attention on the economic effects that arise when a noncompetitive sector exists along with a competitive sector. The recent neoclassical literature on transactions costs and other frictions in a competitive economy could be built into our competitive sector. Indeed, it is suggested below that any lags of adjustment that may occur will be additive to the institutional lags of the noncompetitive sector.[20]

Assuming that the aggregate labor supply is fixed, the supply to the competitive sector may be written as a residual and is given by

$$\bar{L} - S^a\left(\frac{w^a}{w^c}\right) \tag{3}$$

where $\bar{L}$ is the (assumed) constant aggregate supply of labor, and $S^a(w^a/w^c)$ is the labor supply to the administered sector written as a function of relative wages. The derived demand for labor in the competitive sector is assumed to be proportional to aggregate demand, denoted m. On the assumption that the competitive labor market is continually in equilibrium we have

$$km = \bar{L} - S^a\left(\frac{w^a}{w^c}\right) \tag{4}$$

where $k>0$ is the constant proportionality factor in the labor demand equation.[21] Equation (4) indicates that the level of aggregate demand is uniquely related to the relative wage.

maindor of the paper. The qualitative results of later sections do not depend upon this assumption to any significant extent.

[19] As in the section on the administered firm, we assume that the capital stock, technology, and the real wage are fixed.

[20] See, for example, Charles Holt and Dale Mortensen for a discussion of the labor market mechanism in a competitive economy.

[21] In a model with two broad classes of output it is reasonable that the output of both sectors should be strongly superior goods in the sense that a rise in the level of real balances would require an extreme price shift to prevent aggregate demand for the output of either sector from rising. Furthermore, if one conceives of the noncompetitive sector as being composed primarily of manufacturing industries and the competitive sector as being composed primarily of service industries, then the cross-price elasticity between the two sectors is likely to be close to zero.

III. The Macromodel

In this section we will put together the pieces constructed in Sections I and II into a single macromodel. Since the competitive sector has been developed in aggregate form, it only remains to aggregate the administered firms into a single administered sector. We assume that the administered firms all have planning periods of identical length but that the initial point is different for each. If the administered firms have initial decision points distributed uniformly over the interval between T−1 and T, and if all firms are identical, then the administered wage and price indices may be represented by

$$w^a(\mathrm{T}) = \int_{\mathrm{T}-1}^{\mathrm{T}} w_\tau^a d\tau \tag{5a}$$

and

$$p^a(\mathrm{T}) = \int_{\mathrm{T}-1}^{\mathrm{T}} p_\tau^a d\tau \tag{5b}$$

The principal aggregation problem is to find an expression for aggregate unemployment in the economy as a function of the demand proxy variable, m, and the wage and price variables. Since we have assumed that the competitive sector labor market clears continuously, all of the system's unemployment is found in the labor queues of the administered sector. This is, of course, not the only type of unemployment in the economy, and we are abstracting away from the other forms in order to concentrate on the implications of our two-sector labor market model.

The worker can either work in the competitive sector or seek work in the administered sector. The recent literature on job search strongly suggests that the workings of the labor market and the time constraint of a 24-hour day makes it unlikely that a worker can simultaneously search for a job in the administered sector while working in the competitive sector. At any rate, to do so would be costly and this is really all that is required for our analysis.[22]

Figure 2 displays the aggregate demand and supply curves for labor in the administered sector as a function of the relative wage for a given level of aggregate demand, m. As with the product demand in the competitive sector, we assume away relative price effects (due to the level of aggregation), and, given homogeneity, this allows us to write the sector's demand for labor as $D^a(p^a/p^c,\ m)=\beta m$. The curve $S_f^a(w^a/w^c)$ is the full employment supply function that gives the number of workers who would be willing to work in the administered sector if the probability of finding employment in the sector were equal to unity. For $w^a/w^c>1$, it asymptotically approaches the aggregate labor supply $\bar{L}$. It is not coincident with $\bar{L}$ because of the mobility costs associated with moving between sectors. For a relative wage equal to $(w^a/w^c)_0$, the amount of labor hired in the administered sector is $D^a=\beta m$. The horizontal distance between D^a and S_f^c at $(w^a/w^c)_0$, however, does not give the unemployment in the system; as long as unemployment is positive (and thus the probability of find-

[22] The general view in the search literature is that a worker cannot apply once-and-for-all for a job with a firm in the high wage sector and then, while waiting to be called for employment, hold a job in the low wage sector. Hirings are made from the pool of active job applicants; that is, when a vacancy develops, the first "qualified" individual to come into the personnel office is hired. This is less costly to the firm than attempting to trace back old applicants who may have applied first for the job. Indeed, the older the date of application, the less likely it is that the applicant can be found or will still be interested. Consequently, it is the individual who spends his day visiting personnel offices, who is likely to be hired in the high wage sector. In fact, the worker is not constrained to join the queue of a single administered firm; rather, each worker can search among a number of firms. As a consequence, the notion of unemployment is well defined only for the administered sector as a whole and not for individual administered firms.

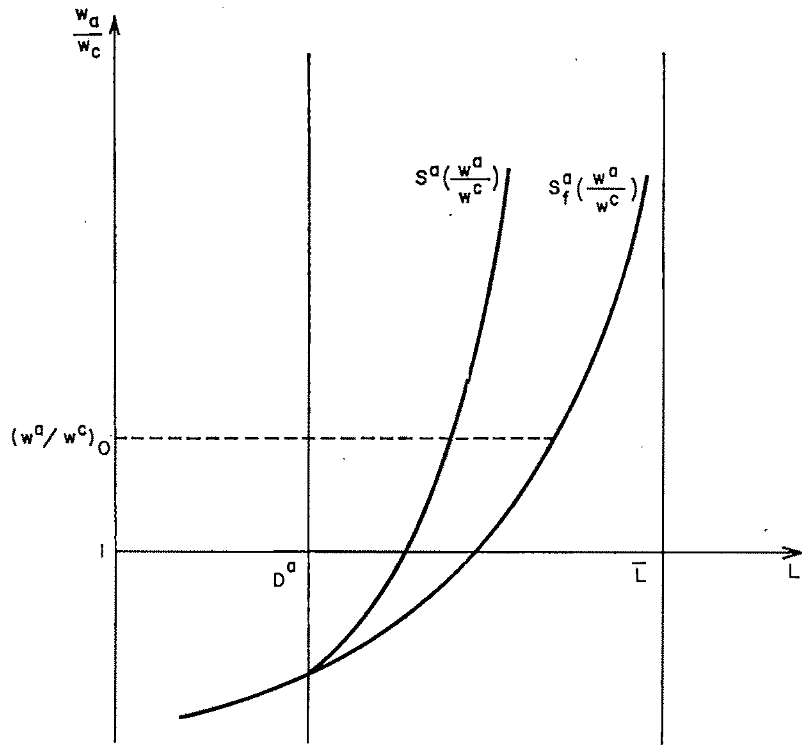


FIGURE 2

ing work is less than unity), the actual supply function to the administered sector is to the left of S^a_f.

The position of the observed (disequilibrium) supply function to the administered sector S^a depends upon two conflicting forces. Namely, as relative wages are increased, the reward for choosing the administered sector rises, but to the extent to which the ratio of S^a/D^a rises, so does the probability of not being employed. The observed supply function is bounded on the right by S^a_f, where the discouraged worker effect is zero, and on the left by D^a, since on D^a the probability of being hired is equal to unity.[23] Thus, S^a will be determined as the solution to an equation of the form:

$$S^a = S^a\left[\frac{w^a(\mathrm{T})}{w^c_{\mathrm{T}}}, \frac{U}{D^a}\right] \tag{6}$$

where $U = S^a - D^a$ is the amount of unemployment in the system.[24] Assuming

[23] The supply function S^a is observable only where its slope is greater than the slope of the demand function. If this were not the case, unemployment would fall as the relative wage was increased, and to the extent that the administered sector was aware of the shape of S^a, it would never increase its relative wage into this region. Notice that if $D^a \to 0$ as the relative wage rose then S^a could actually fall with rising relative wages. The workings of the supply mechanism with demand shifts are obvious. An increase in real balances shifts the administered sector demand function to the right, and if relative wages are held constant, this will increase employment in the sector. The increase in employment increases the probability of being hired and this, in turn, shifts the supply function S^a towards the full employment supply S^f.

[24] In this model the existing unemployment is not strictly search unemployment in the sense that we do not have heterogeneous workers searching among heterogeneous jobs and there is no lack of information as to the probability of a worker being hired in the ad-

that this equation may be solved explicitly for U and approximated by a linear form, we have

$$(7) \qquad U \cong e\frac{w^a(\mathrm{T})}{w^c_{\mathrm{T}}} + f - hD^a$$
$$= e\frac{w^a(\mathrm{T})}{w^c_{\mathrm{T}}} + f - h\beta m$$

where $e, f > 0$, and $0 < h < 1$. The parameter k may differ from unity because of the influence of the change in the probability of being hired upon the supply of labor. For simplicity, we will assume that h is constant.

The unemployment described by equation (7) is similar, in a theoretical sense, to the unemployment that arises in a dual labor market setting which contains a high wage sector and a relatively low wage, market clearing sector. Although the dual labor market has traditionally been viewed as relevant to underdeveloped economies, current empirical research suggests that it is also descriptive of developed economies such as that of the United States.[25] As mentioned above, however, this paper assumes market clearing in the competitive sector solely to concentrate on the effects introduced by the interaction between the two sectors. The effects of this interaction on wage determination and inflation, however, are not sensitive to the precise assumptions concerning market clearing.

For reference, the complete macromodel is written below. The wage and price equations are

$$(5a) \qquad w^c(\mathrm{T}) \equiv \int_{\mathrm{T}-1}^{\mathrm{T}} w^a_\tau d\tau$$

and

$$(5b) \qquad p^c(\mathrm{T}) \equiv \int_{\mathrm{T}-1}^{\mathrm{T}} p^a_\tau d\tau$$

where the administered markups are given by

$$(2a) \qquad w^a_\tau = \phi(m)\Psi E_\tau w^c_\tau$$

and

$$(2b) \qquad p^a_\tau = \mu(m)\Psi E_\tau p^c_\tau$$

Aggregate unemployment is given by

$$(7) \qquad U = e\frac{w^a(\mathrm{T})}{w^c_{\mathrm{T}}} + f - h\beta m$$

Using (7), the competitive market clearing equation may be written as

$$(4') \qquad cm = \left\{\bar{L} - \left(e\frac{w^a(\mathrm{T})}{w^c_{\mathrm{T}}} + f\right)\right\}$$

where

$$c \equiv k + (1-h)\beta > 0$$

Equation (4′) implies that the relative wage structure is a function of aggregate demand alone:

$$(8) \qquad \frac{w^a(\mathrm{T})}{w^c_{\mathrm{T}}} = \frac{(\bar{L} - f) - cm}{e}$$

Hence, unemployment becomes a function of aggregate demand, i.e.,

$$(9) \qquad U = e\frac{w^a(\mathrm{T})}{w^c_{\mathrm{T}}} + f - h\beta m$$

ministered sector. The assumptions of homogeneous workers and jobs and perfect information on the probability of being hired are made only to simplify the analysis. The model, however, may be extended—without altering the conclusions of the paper—by explicitly introducing search considerations.

[25] A component of unemployment in the dual theories arises because the price of leisure is greater than the wage in the low wage sector, but lower than the wage in the high wage sector. Hence, workers who are unemployed for a prolonged period eventually leave the labor force. In this paper, the size of the labor force is assumed to be exogenous so that unemployed workers are assumed to prefer work in the low wage sector than to be out of the labor force. This assumption, which is made to simplify the analysis, can easily be altered. For a still more general theory one could also introduce heterogeneous workers and jobs as discussed in fn. 24.

$$= e\left[\frac{(\bar{L} - f) - cm}{e}\right] + f - h\beta m$$
$$= \bar{L} - (c + h\beta)m$$

IV. Steady-State Dynamics

Constant Expectations

In this section we investigate the manner in which the two sectors interact in the macromodel. We shall find that the steady-state dynamics of the model can describe many empirically observed phenomena. It is easiest to begin by analyzing the case where expectations are constant, $E_t = E$, and where competitive wages and hence prices are growing at a constant rate of growth ξ, i.e.,

$$(10) \qquad Dw_T^c = Dp_T^c = \xi$$

where $Dx \equiv (1/x)dx/d\tau$. From the definitions of the aggregate wages and prices (5) and the markup equations (2) we have

$$(11) \quad w^a(\mathrm{T}) = \int_{\mathrm{T}-1}^{\mathrm{T}} w_\tau^a d\tau = \int_{\mathrm{T}-1}^{\mathrm{T}} \phi\Psi E w_\tau^c d\tau$$
$$= \phi\Psi E g(\xi) w_t^c$$

where

$$g(\xi) \equiv \int_{\mathrm{T}-1}^{\mathrm{T}} e^{-\xi(\mathrm{T}-\tau)} d\tau = \frac{1}{\xi}(1 - e^{-\xi})$$

Hence, relative wages and prices are given by

$$(12a) \qquad \frac{w^a(\mathrm{T})}{w_\mathrm{T}^c} = \phi(m)\Psi E g(\xi)$$

and

$$(12b) \qquad \frac{p^a(\mathrm{T})}{p_t^c} = \mu(m)\Psi E g(\xi)$$

Since $g(\xi)$, the growth markup, declines monotonically in ξ, the wage and price premium of the administered sector varies inversely with the rate of change of wages and prices. To understand this result, suppose that the system is originally in a dynamic equilibrium with wages and prices increasing at the rate ξ in both sectors. Let the system be disturbed by an increase in Dw_T^c to $\xi' > \xi$. Since the entire competitive sector adjusts continuously, ξ' is the rate of wage change for all competitive firms. The administered sector, however, adjusts with a lag; the only administered firms that can adjust immediately to the change are those at the beginning of a new planning period. The result, in steady-state equilibrium, is a narrowing of the wage structure. Furthermore, as long as expectations are unchanged (an assumption that we will drop in the next section), the wage structure will remain indefinitely in the narrowed position given by the equilibrium equations (12a) and (12b). Although the representative administered firm "catches up" with the competitive sector at the beginning of its planning period, the firm plans only an inflation markup of E for the planning period. Consequently, the larger is ξ', given E, the further behind the firm will be at the end of the planning period. (This does not mean that $w^a(\mathrm{T})$ necessarily falls behind w_T^c. That result would depend upon the difference between Dw_T^c and E_T in relation to the markup $\phi\Psi$.) Since the economy, at any point in time, consists of administered firms at different stages of their respective planning periods, it follows that the greater the growth rate, the narrower the wage structure.[26]

The above analysis is useful in tracing out the effects of changes in aggregate demand or real balances on the rate of inflation and unemployment, as the economy moves across steady states with

[26] For a further discussion of the mechanism through which the wage structure varies over the cycle, see Wachter (1972). Evidence that the wage structure narrows when aggregate demand rises (and thus unemployment declines) also may be found in the Wachter paper.

constant expectations. Given our assumptions about price elasticities, an increase in m appears as an increase in demand to both sectors. For the high wage sector the immediate adjustment means an increase in sector employment and thus a decrease in the size of the queue. If the competitive sector also is to meet the increase in demand, it must attract labor from the queue of the high wage sector. As a result, it must increase its relative wage. Differentiating (8) with respect to m, we obtain

$$\frac{\partial(w^a(\mathrm{T})/w^c_{\mathrm{T}})}{\partial m} = -c/e < 0$$

that is, to allow increased hiring in the competitive sector, the wage structure narrows with an increase in real demand. Clearly, however, this narrowing cannot be brought about simply by a once-and-for-all increase in the competitive wage level since such a wage increase would be matched by firms in the administered sector. The only way for the low wage firms to succeed in raising their relative wage is to increase their *rate* of wage change Dw^c_{T}. The mechanism through which an increase in Dw^c_{T} or ξ narrows the wage structure was described above.

To formally verify that an increase in demand increases the steady-state rate of inflation, we differentiate (12a) with respect to m,

$$(13)\quad \Psi E\left(\phi'(m)g(\xi) + \phi(m)g'(\xi)\frac{\partial \xi}{\partial m}\right) = \frac{\partial}{\partial m}\left(\frac{w^a(\mathrm{T})}{w^c_{\mathrm{T}}}\right) = -\frac{c}{e} < 0$$

and solve for

$$(14)\quad \frac{\partial \xi}{\partial m} = \frac{-\frac{c}{e} - \Psi E\phi'(m)g(\xi)}{\Psi E\phi(m)g'(\xi)}$$

Since the markup increases with increases in demand, $\phi'(m)>0$, and since $g'(\xi)<0$, we have $\partial\Psi/\partial m>0$.[27]

The impact of changes in m on unemployment can be obtained by differentiating (9) with respect to m. It is obvious that an increase in real balances or aggregate demand causes a decrease in unemployment.

The implications of an increase in aggregate demand may be summarized in the following manner. Assume the system is originally in equilibrium with $\xi=E$. For both sectors the increase in aggregate demand causes a shift in the derived demand for labor. The high wage sector can satisfy its increased demand by drawing upon its labor queue, that is, reducing unemployment without increasing its rate of wage change. The low wage competitive sector, on the other hand, needs to increase its wage offer relative to the high wage sector in order to attract the additional workers from the unemployment pool of the high wage sector. The increase in relative wages is achieved for constant expectations by increasing Dw^c_{T}. Furthermore, the larger the increase in aggregate demand, the greater the increase in the rate of inflation, the narrower the wage structure, and hence, the smaller the remaining unemployment.[28] This inverse relationship between inflation and unemployment is the short-run Phillips curve displayed in Figure 3.[29]

Realized Expectations

The Phillips curve is a steady-state concept, drawn for a particular level of expectations. Friedman and Phelps have

[27] The assertion that $\phi'(m)>0$ is an intuitively clear result and will be assumed below.

[28] Notice that since a given wage structure is associated with a given unemployment rate, a low unemployment rate means a high rate of aggregate wage and price change, but not a constantly changing relative wage.

[29] The convexity of the Phillips curve is assumed.

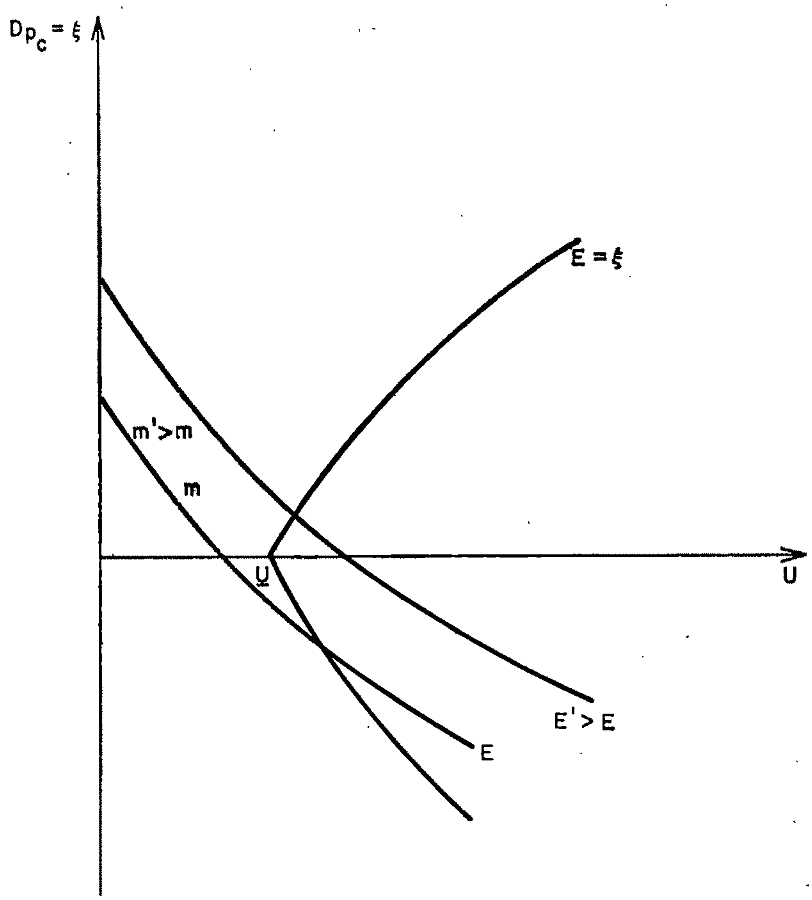


Figure 3

argued that the curve will shift vertically upward by 1 percent for every 1 percent increase in expectations. Suppose on the Phillips curve of Figure 3 the point (U, ξ) had the property that expectations were realized, i.e., the actual rate of inflation ξ was the rate expected. If a rise of $\Delta\xi$ in expectations caused the curve to shift upward by $\Delta\xi$, then $(U, \xi+\Delta\xi)$ would be the point on the new curve where expectations were realized, and there would have been no alteration in unemployment, U (hence, m could remain constant). In other words, the fully anticipated inflation has had no real effects, and if we connected all the points at U, we would have a vertical long-run Phillips curve. In our model this is not the case and inflation, even when fully anticipated, will have real effects.[30]

Since expectations are realized, in continuous time E_T will be some function of the actual rate, and for the present we let

$$E_T = E\{(p^c_{T+1/2})/p^c_T\}$$
$$= p^c_T/p^c_{T-1/2} = e^{\xi/2}$$

justifying this assumption below. Substituting this into the wage structure equation (12a) we obtain

[30] This is a bit ambiguous. More precisely, the inflation has no real effects if aggregate demand is held constant, but then expectations cannot be realized at all rates of growth. This is discussed below.

$$(15)\qquad \frac{w^a(\mathrm{T})}{w_{\mathrm{T}}^c} = \phi\Psi e^{\xi/2}g(\xi) \equiv \phi\Psi h(\xi)$$

where $h(\xi)\equiv e^{\xi/2}g(\xi)$ is the growth markup with realized expectations. Unlike $g(\xi)$, $h(\xi)$ is an increasing function of the growth rate (for $\xi>0$), and with realized expectations the wage structure widens with increases in the rate of wage and price inflation.[31]

The response of the system to changes in aggregate demand and the nature of the long-run Phillips curve may now be identified. From (14) we have

$$(16)\qquad \frac{\partial\xi}{\partial m} = \frac{-\dfrac{c}{e} - \Psi\phi'(m)h(\xi)}{\Psi\phi(m)h'(\xi)} < 0$$

since $\phi'(m)>0$ and $h'(\xi)>0$. That is, when expectations are realized, an increase in real balances is associated with a lower rate of wage and price inflation. As a result, since an increase in real balances (or aggregate demand) reduces unemployment, the long-run Phillips curve (Figure 3) has a positive slope, i.e.,

$$(17)\qquad \frac{\partial\xi}{\partial U} = \frac{\partial\xi}{\partial m}\frac{\partial m}{\partial U} > 0$$

The emergence of an upward sloping long-run Phillips curve has its roots in the assumption that the administered firm finds a customer queue more costly than a worker queue or, alternatively (where the high wage sector is a unionized sector), that the union desires a wage premium for its members. This implies that the firm sets its wage and price decision so as to have a worker queue over the bulk of the planning period, i.e., the full employment point, γ of Figure 1, is reached near the end of the period. To see heuristically that this implies an upward sloping Phillips curve, suppose that the actual and realized rate of inflation increases. Assume that at the initial rate of inflation the firm would be at point a, in Figure 4, at the end of the planning period, whereas at the higher rate of inflation it would be at point b at T+1. If the firm always desires a worker queue, γ is set at the end of the period and an increase in ξ causes firms to increase their relative wage and thus to alter L^D and L^S so that they intersect at b rather than at a in Figure 4. Thus, if a firm desires to maintain a worker queue throughout the planning period, an increase in the rate of inflation forces the firm to increase its relative wage and hence the average number of workers on the queue (i.e., the number of unemployed workers) throughout the period also increases. It is shown above, algebraically, that if firms look roughly to the middle of the period, i.e., $E=e^{\xi/2}$, this is sufficient to override the growth markup, $g(\xi)$, making $h(\xi)$ positive and hence yielding an upward sloping Phillips curve. Since firms look more to the end than to the beginning of the period, they set $E>e^{\xi/2}$, and the upward sloping curve holds *a fortiori*.

In summary, the markup $g(\xi)$ indicates the extent to which the administered sector lags behind the competitive sector. Consequently, as the rate of inflation, ξ, increases and $g(\xi)$ declines, the wage premium narrows. When the increase in ξ is anticipated, though, administered firms respond by increasing their wage markup. This offsets the narrowing of the premium induced by the lag of the administered sector. The net result is an increase in the wage differential and, hence, in unemployment.

[31] Letting $h(x)=(e^x-e^{-x})/x$, sign $h'(x)=$ sign $\{(x-1)e^x+(x+1)e^{-x}\}$. Since $h'(0)=0$ and $d/dx[(x-1)e^x+(x+1)e^{-x}]=x(e^x-e^{-x})$ for $x>0$, we must have $h'(x)>0$, and for $x<0$, $h'(x)<0$. In what follows we will consider only the case where $\xi>0$. For the alternative case, see the discussion at the end of this section.

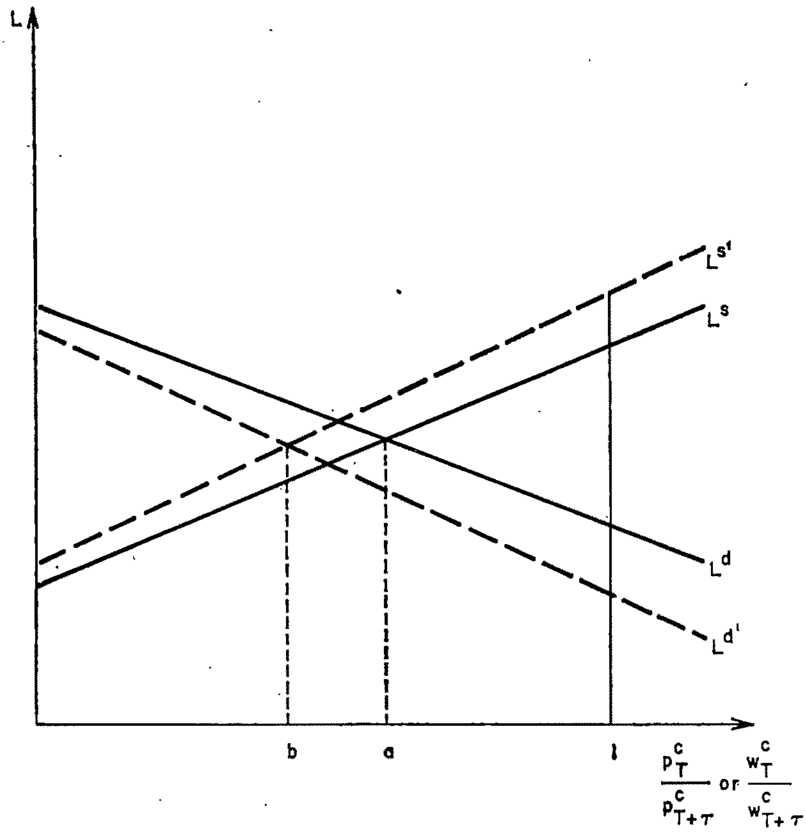


FIGURE 4

In the case where the economy is deflating, the above arguments are reversed. Now the firm will attempt to set the full employment point γ (perhaps to the right of unity) near the beginning of the planning period, allowing the deflation to carry it into a worker queue region over the bulk of the period. Consequently, the deflation markup will be close to unity, and $h(\xi)$ will be close to $g(\xi)$. The signs of (16) and (17) will now be reversed, and the long-run Phillips curve will be downward sloping (see Figure 3). This implies that there is a minimal level of unemployment, U in Figure 3, which can be attained by aggregate demand policies alone when expectations are realized. Attempts to achieve lower unemployment than U through aggregate demand policies, cannot succeed in the long run.[32]

A caveat is necessary in concluding this section. Although we have derived a so-called long-run Phillips curve, the notion of long-run equilibrium in an economy with noncompetitive elements is ambiguous. In particular, it is possible that institutional features may change in the long run in response to economic variables.

[32] Although the c-shaped long-run Phillips curve in the above analysis is based on the assumed attitude of the administered firm towards a customer queue, other reasonable behavioral assumptions could also lead to the same result. For example, if administered firms view demand as being more uncertain when prices and wages are changing rapidly rather than when they are changing slowly, the safety factor, Ψ, would become a function of ξ so as to result in a c-shaped Phillips curve.

For example, the length of the planning period, which is assumed to be fixed in the previous analysis, may shorten as a long-run response to a rapid (positive or negative) rate of inflation. Escalator clauses would have this effect. A model where the length of the planning period is an endogenous variable is presented in our 1972 discussion paper.

V. Theoretical and Policy Implications

The Phillips Curve

The model presented in this paper suggests several labor market features that were largely ignored in earlier work on the Phillips curve. The stability of expectations is a function of the time horizon of the forecast. In most empirical work that takes account of expectations, the time horizon of the expectation is anywhere from one quarter to one year; that is, the market is viewed as being concerned with wage and price changes (or wage and price levels) that are expected to occur within the next year. The presence of a planning period lag, however, implies that the time horizon for expectations should at least be equal to the length of the planning period. In certain key manufacturing sectors of the economy the presence of union-negotiated contracts creates a planning period of two or three years in duration, which implies a relatively long time horizon. As the expectation time horizon increases in length, the expectations series for a cyclical economy such as that of the United States becomes smoother; that is, whereas the expected rate of price change anticipated by businessmen for, say, 1972 may be large (in relation to the historical mean), the expected rate of price change for the longer period of 1972 through 1975 is likely to be considerably closer to the historical mean. The smaller is the variance of the expectations series, of course, the more stable is the Phillips curve. Only significant changes in expectations or changes in expectations that do not average out over a three-year period appear in the planning period expectational series and cause a shift in the Phillips curve.[33]

Although the expectational effect increases the stability of the Phillips curve, it also introduces a long lag between current wage and price changes and past rates of inflation. Furthermore, this lag is augmented by the catch-up effect; that is, a firm which guesses incorrectly the developments over its contract period, must wait at least until the beginning of a new planning period before it can rectify the past error. The expectational and catch-up effects operate in the same direction. When the current inflation rate is, for example, above the expected rate, not only are expectations likely to increase, but also the catch-up is positive. A lack of attention to these long lags may help to explain the tendency of empirical Phillips curves to underpredict the rate of inflation between 1969 and 1972.

Social Costs and Stability

The social cost of having an administered sector is associated with the unemployment on the labor queue of the high wage sector.[34] The cost of unemployment

[33] The stability of the short-run Phillips curve is further enhanced by the inclusion of short-run frictions such as money illusion which will interact with the planning period effect in an additive fashion. Given an unanticipated inflation, the presence of money illusion on the part of workers, for example, may be indicated by an increase in the aggregate labor supply $\bar{L}$ at any given real wage rate. This enables the low wage sector to attract additional workers without the necessity of increasing relative wages (in order to draw down the queue of the high wage sector). In turn, to the extent that the queue to the high wage sector is not reduced, there is no reason for that sector to revise its wage offer. As the money illusion wears off so that $\bar{L}$ returns to its former level, the low wage sector is forced to increase ξ in order to attract additional workers. It is only at this point, with the increase in ξ and the decrease in the labor queue, that the high wage sector need adjust its wage offer.

[34] In fact, since we assume that the low wage sector clears continuously, the only cause of unemployment

is a cost to society in the form of output foregone; that is, there is a production inefficiency (arising in part from the divergence in wage rates). In addition to efficiency considerations, there is also a redistribution of income to those who are employed in the high wage sector from those who are not (whether they are in the queue or in the competitive sector).

The dynamic costs to society imposed by the administered sector turn on the question of macro-economic stabilization policy.[35] To the extent that it is understood by the policy makers, the lags associated with the planning period effect increase the stability of the system. If the dynamics of the mechanism are misunderstood, for example, if the government interprets the short-run Phillips curve as a loci of attainable long-run options, attempts to control the economy can be destabilizing.

One important potential for error concerns the identification of the source of the inflation. For example, suppose that the government increases aggregate demand from m to m' (in Figure 3) but then becomes dissatisfied with the resulting rate of inflation ξ'. It, therefore, decreases demand back to m. Unemployment increases back to U, but since administered firms are still adjusting to the past price increases, the rate of inflation at first remains above ξ. Due to the planning period lags, the government cannot lower the rate of inflation as fast as it can increase the unemployment rate.[36] The danger is that the government may erroneously decide that the economy is suffering from a cost-push inflation. Furthermore, the evidence would lend misleading support to this view. Despite the fact that unemployment is rising, the rate of wage change in the high wage (possibly unionized) sector would be greater than the wage change in the competitive sector, widening the wage structure and giving (to what is a catch-up phenomenon) the appearance of a wage or cost-push. This exercise is meant as a capsule summary of the recent *U.S.* inflation experience in the period from 1968–71, culminating in the adoption of wage-price controls in August 1971. For a further discussion of this point see the paper by Wachter (1972).

A particularly intractable problem arises when the administered sector has expectations of inflation that are greater than the rate of inflation which is desired by the government. The problem in this case is that the administered sector firms lock themselves into the higher expected rate for the length of the planning period, and the government then faces the choice of either lowering aggregate demand in the hope of reducing expectations for those firms that are beginning their planning period or validating the higher rate of inflation by increasing real balances. In the former case the anti-inflation policy results in both high unemployment and high inflation as long as it takes for expectations to be adjusted downward. If the latter policy is adopted, low unemployment can be maintained but only at the cost of spiraling inflation.

To summarize, the planning period effect permits the government to reverse any shocks that may increase the rate of inflation to an undesirable level, but if the government misinterprets the lags in the structure, it may adopt policies which are themselves destabilizing. In addition, differences between the expectations of the administered sector and the government's desired rate of inflation will be particularly difficult to deal with.

in the model is due to the queue for the administered firms. It should be stressed again that our assumption of continuous market clearing in the competitive sector is made solely to permit us to concentrate on the inflation effects of assuming a two-sector market.

[35] We will be concerned in this section principally with the instability issue itself and will not attempt to formally link this to social costs.

[36] This result may be derived from the disequilibrium dynamics of the model.

REFERENCES

G. A. Akerlof, "Relative Wages and the Rate of Inflation," *Quart. J. Econ.*, Aug. 1969, *83*, 353–75.

R. J. Barro and H. J. Grossman, "A General Disequilibrium Model of Income and Employment," *Amer. Econ. Rev.*, Mar. 1971, *61*, 82–93.

R. Clower, "The Keynesian Counter-Revolution: A Theoretical Appraisal," in F. M. Hahn and F. P. R. Brechling, eds., *The Theory of Interest Rates*, London 1965.

P. B. Doeringer and M. J. Piore, *Internal Labor Markets and Manpower Analysis*, Lexington 1971.

J. T. Dunlop, *Wage Determination Under Trade Unions*, New York 1944.

———, "The Task of Contemporary Wage Theory," in G. Taylor and F. Pierson, eds., *New Concepts in Wage Determination*, New York 1957.

O. Eckstein and G. Fromm, "The Price Equation," *Amer. Econ. Rev.*, Dec. 1968, *58*, 1158–83.

O. Eckstein and T. Wilson, "The Determinants of Money Wages in American Industry," *Quart. J. Econ.*, Aug. 1962, *70*, 379–414.

M. Friedman, "The Role of Monetary Policy," *Amer. Econ. Rev.*, Mar. 1968, *58*, 1–17.

J. K. Galbraith, *The New Industrial State*, Boston 1967.

R. J. Gordon, "Problems in Predicting the Rate of Inflation," paper presented at the North American Regional Conference of the Econometric Society, New York, Dec. 1969.

J. R. Harris and M. P. Todaro, "Migration, Unemployment and Development: A Two-Sector Analysis," *Amer. Econ. Rev.*, Mar. 1970, *60*, 126–42.

B. Harrison, *Education, Training and the Urban Ghetto*, Baltimore 1972.

C. C. Holt, "Job Search, Phillips' Wage Relation, and Union Influence: Theory and Evidence," in E. S. Phelps et. al. eds., *Macroeconomic Foundations of Employment and Inflation Theory*, New York 1970.

W. A. Lewis, "Economic Development with Unlimited Supplies of Labour," *Manchester Sch. Econ. Soc. Stud.*, May 1954, *22*, 139–91.

R. F. Lucas and L. A. Rapping, "Real Wages, Employment, and Inflation," *J. Polit. Econ.*, Sept. 1969, *77*, 721–54.

D. T. Mortensen, "Job Search, The Duration of Unemployment and the Phillips Curve," *Amer. Econ. Rev.*, Dec. 1970, *60*, 847–62.

D. Patinkin, *Money, Interest, and Prices*, 2d ed., New York 1965.

G. L. Perry, *Unemployment, Money Wage Rates, and Inflation*, Cambridge 1966.

E. S. Phelps, "Money Wage Dynamics and Labor Market Equilibrium," in E. S. Phelps et. al., eds., *Microeconomic Foundations of Employment and Inflation Theory*, New York 1970.

S. A. Ross and M. L. Wachter, "The Pricing and Timing Decision of the Oligopoly Firm," Univ. Pennsylvania disc. pap., Sept. 1972.

F. M. Scherer, *Industrial Market Structure and Economic Performance*, Chicago 1970.

R. M. Solow, "Recent Controversy on the Theory of Inflation: An Eclectic View," in S. W. Rousseas, ed., *Inflation: Its Causes, Consequences, and Control*, New York 1968.

G. J. Stigler and J. K. Kindahl, *The Behavior of Industrial Prices*, New York 1970.

M. P. Todaro, "A Model of Labor Migration and Urban Unemployment in Less Developed Countries," *Amer. Econ. Rev.*, Mar. 1969, *59*, 138–48.

M. L. Wachter, "Cyclical Variation in the Interindustry Wage Structure," *Amer. Econ. Rev.*, Mar. 1970, *60*, 75–84.

———, "Phase II, Cost-Push Inflation, and Relative Wages," mimeo. 1972.

E. R. Weintraub and S. Weintraub, "The Full-Employment Model: A Critique," *Kyklos*, 1972, *25*, 83–100.

S. Weintraub, *Some Aspects of Wage Theory and Policy*, Philadelphia 1963.

Male-Female Pay Differentials in Professional Employment

By BURTON G. MALKIEL AND JUDITH A. MALKIEL*

This study analyzes salary differentials among a sample of 272 professional employees of a single corporation. While it is dangerous to generalize from the practices of one employer, a micro-economic case study approach has several advantages. Firstly, because the employer in question hires large numbers of men and women doing the same range of jobs, a measure of the extent of sex discrimination can be obtained, holding occupation constant. It has been almost impossible to obtain such estimates in the past, since as Victor Fuchs has indicated, men are seldom employed in the occupations employing large numbers of women, and vice versa. Secondly, because the study has been confined to highly educated professional employees, the sample of men and women is relatively homogeneous with respect to career interests and attachment to the labor force. Other studies of sex discrimination have usually failed to distinguish between two very different groups of women employees: those with a long-run commitment to their occupation and those whose attachment to the labor force is temporary. Thirdly, since the employer opened its full personnel records to examination, it was possible to obtain detailed information on the previous experience and personal characteristics influencing salary differentials.[1] Such information has generally been unavailable in previous studies. Finally, because the employees in the sample have invested considerable years in higher education and in postschooling investments, an analysis of the salary differentials of this sample should provide information on the relevance of recent explanatory models of wage differentials in the human capital tradition at a level of disaggregation never used before.

After a brief review of the postschooling investment model as an explanation of salary differentials, estimates of the model as applied to men and women employees separately are presented in Section II. A somewhat expanded model, including productivity measures relevant to the employing organization, is estimated in Section III. In Section IV, the concept of discrimination is examined and estimates of discrimination against women are presented. In Section V, a "job level" variable is added to the equations and the assignment of men and women to job levels is analyzed. We conclude in Section VI that the source of any discrimination against women is in job assignments rather than in pay differentials between the sexes for the same job.

I. The Postschooling Investment Model

According to the human capital approach developed by Gary Becker and Jacob Mincer, an individual's gross an-

* Professor of economics, Princeton University and examiner, Educational Testing Service. We are grateful to Orley Ashenfelter, Alan Blinder, George Borts, Daniel Hamermesh, Jacob Mincer, Richard Quandt, and Albert Rees for many helpful comments and suggestions; Ellen Yaroshevsky for collecting the data; Elvira F. Giaimo for programming the calculations; and the National Science Foundation for financial support.

[1] We are extremely grateful to the organization, which wished to remain anonymous, for its generosity in providing us with all the personnel information available to it.

nual earnings can be related to past investments as follows:

$$W_t = W_0 + r \sum_{i=1}^{t-1} C_i \tag{1}$$

where W_t is the gross annual wage in period t; C_i is the cost of human capital investment during period i; r is the rate of return (assumed to be constant over time) to human capital investment; and W_0 is the annual wage that would be received in the absence of any human capital investments.

If formal education is the only capital investment and if the costs can be estimated as the earnings foregone during schooling, Becker and Barry Chiswick as well as Mincer have shown that equation (1) may be rewritten as

$$ln\,(W_t) = ln\,(W_0) + rS \tag{2}$$

where S represents the number of years of formal education in which the individual has specialized in human capital formation. The model can also be expanded to include investment in on-the-job training. Letting k_0 stand for the fraction of time invested in on-the-job training at the start of work experience, and assuming that such investments decline linearly over time (as by the relationship $k_j = k_0 - (k_0/T)_j$, where T stands for the years of work experience when on-the-job training ends), we may modify equation (2) as follows:

$$ln\,(W_t) = ln\,(W_0) + r_s S + r_e k_0 j - r_e \frac{k_0}{2T} j^2 \tag{3}$$

Such formulations may be found in Mincer and Thomas Johnson.

In (3), the subscripts s and e allow the rates of return for schooling and experience to differ, and j stands for the number of years of work experience. The *log* of wages is related linearly to the years of schooling (S) and in a quadratic fashion to the years of work experience (j). We call this the narrow form of the postschooling investment model.

II. Estimation of the Narrow Postschooling Investment Model

This section presents cross-sectional estimates of the narrow form of the model. The dependent variable was the employee's annual salary (or the full-time equivalent annual salary for the few part-time workers in the sample).[2] The number of years of schooling was measured as the years of education beyond high school. It was necessary in many cases to divide some of the years of graduate training into partial years of schooling and partial years of experience, as, for example, when the student held a part-time research or teaching job. The detail in the individual personnel folders made such divisions possible. Experience (outside of the present organization) was broken down into two categories: related (to the present job) work experience and unrelated experience. An example will illustrate how the division was made.

Suppose a man with two years of experience at a university or research organization was hired as a research scientist. Clearly his experience is related to the present job and undoubtedly considerable on-the-job training investments were undertaken during that period. On the other hand, if the individual came to work as a research scientist with two years' experience as a busboy after school-

[2] This wage is net of any postschooling investment in the current year. According to equations (1) to (3), however, it is the gross wage (before investment) rather than the net wage paid that is the correct dependent variable. It is likely that the discrepancy is very small, however, especially for the year 1971, since only workers already in the organization by 1968 are included in the 1971 sample. Presumably these employees have completed a considerable part of their apprenticeship period during which most of the postschooling investment takes place.

TABLE 1—ESTIMATES OF NARROW FORM OF THE POSTSCHOOLING INVESTMENT MODEL[a,b]

1969

Sample Group	Constant Term	Coefficient S	Coefficient j	Coefficient j^2	R^2	F-Value	Standard Error of Estimate
Men	8.697	+0.081 (9.93)	+0.054 (11.37)	−0.001 (8.17)	0.71	125.07 (3,155)	0.167
Women	8.504	+0.066 (6.47)	+0.059 (8.91)	−0.001 (5.45)	0.68	78.89 (3,109)	0.192

[a] Numbers in parentheses below the coefficients are t-values.
[b] S is years of schooling after high school, j is years of seniority and relevant work experience. Underneath the F-value the degrees of freedom are shown; the standard error of estimate is adjusted for degrees of freedom.

ing, the experience was coded as unrelated. Experience was coded as unrelated only if it was clearly irrelevant to the present occupation. Experience in the present organization (seniority) was simply the number of years employed. As will be explained below, related work experience outside the organization and work experience within the organization were combined into one experience variable.

The sample consists of 272 employees (159 men and 113 women) employed in fiscal year 1969 in six divisions. Their work was considered technical and/or scientific and some degree of advanced training was usually a requirement. Salaries and other personnel data were available backward in time to 1966 and forward in time through 1971. The sample numbers, however, decline in both directions (particularly going backward in time) since some of the employees at work in 1969 were not employed in 1966 and some of the 1969 employees subsequently left the organization or retired.

Cross-sectional estimates of equation (3), the narrow form of the postschooling investment model, for 1969, are presented in Table 1. As equation (3) shows, the coefficient of S may be interpreted as the rate of return to schooling. According to our estimates, men received a rate of return of 8.1 percent while women received only 6.6 percent.[3] The difference was significant at the .05 level. While the rate of return to on-the-job training (r_e) is not identified, our estimates suggest that it is at least as high as that attributable to schooling. For example, assuming that k_0, the fraction of time invested in on-the-job training at the start of work experience is one-half, one obtains an estimate of r_e in the men's equation of over 10 percent, since the coefficient of j is the product of $r_e k_0$. The number of years during which on-the-job training takes place, T, may be estimated as one-half the ratio of the coefficients of j (the experience variable) and j^2. For men the estimate is 27 years, a fairly long period, but not unreasonable in view of the technical nature of the work involved.

In Table 1, related outside work experience and experience within the organization (seniority) were combined into one experience variable. The estimated coefficients of the two types of on-the-job training were very similar when they were

[3] Equation (3) was also estimated under the assumption that the schooling variable might enter the equation as a quadratic. The estimated coefficient of the quadratic term was always negative, suggesting diminishing returns to schooling. Nevertheless, the standard error of estimate for the equation, adjusted for degrees of freedom, was never improved by the quadratic specification and therefore the linear specification was retained in this study.

TABLE 2—ESTIMATES OF NARROW FORM OF THE POSTSCHOOLING INVESTMENT MODEL USING APPROXIMATION FOR WORK EXPERIENCE[a,b]
1969

Sample Group	Constant Term	Coefficient S	Coefficient j'	Coefficient j'^2	R^2	F-Values	Standard Error of Estimate
Men	8.674	+0.091 (10.62)	+0.044 (8.83)	−0.001 (5.76)	0.66	100.63 (3,155)	0.180
Women	8.573	+0.078 (6.00)	+0.037 (5.32)	−0.001 (4.07)	0.47	32.07 (3,109)	0.248

[a] Numbers in parentheses below the coefficients are t-values.
[b] $j' = Age - (18+S)$. Other terms are defined in Table 1.

allowed to enter the equation separately.[4] Nonrelated work experience was dropped from the equation since estimates of the coefficient of such experience were never significant.[5] As will be indicated later, we also were able to reject the hypothesis that the salaries of men and women were generated by the same pay structure.

Most previous empirical studies have not had available precise information on work experience. Such experience has usually been estimated as the number of years since the completion of schooling, i.e., $j' = Age - (18+S)$. The results of using such an estimate of work experience is presented in Table 2.

Comparison of the two tables indicates that the crude approximation of experience seems to work well for the sample of men. The R^2 and the coefficient estimates are quite similar in the two tables. The differences are much greater for the women's equations and the R^2 falls sharply. The approximation worked less well for women since many women in the sample spent several years after formal schooling out of the labor force or in nonrelated work activities.

We also attempted to add age variables to equation (3), which included only schooling and experience. Perhaps there is a maturation process with age that affects salaries apart from the postschooling investments associated with work experience. To answer this question we estimated equation (4) but could not accept the hypothesis that the coefficients of the added age variables were jointly significantly different from zero.[6]

[4] For example, during 1969, the following coefficients (and t-values) were obtained when seniority and related work experience were allowed to enter equation (3) separately, as first and second degree terms.

	Men	Women
Related Work Experience	+ 0.041 (6.00)	+ 0.044 (4.46)
Related Work Experience²	− 0.001 (3.85)	− 0.001 (2.09)
Seniority	+ 0.049 (4.64)	+ 0.046 (3.47)
Seniority²	− 0.001 (2.92)	− 0.001 (1.65)

It should be also noted that the standard error of estimate (adjusted for degrees of freedom) for the equation was higher when seniority and related experience entered the equations separately.

[5] For example, for the year 1969 when the number of years of nonrelated work experience was added (as a separate independent variable) to equation (3), the estimated coefficients for men and women were +0.003 with standard errors of (0.004).

[6] The hypothesis was tested by calculating the ratio

$$F = \frac{[R^2(4) - R^2(3)]/2}{[1 - R^2(4)]/(n-6)}$$

where $R^2(4)$ and $R^2(3)$ are the coefficients of determination obtained from the estimates of equations (4) and (3), respectively, and n is the number of observations.

$$(4) \quad ln\,(W_t) = ln\,(W_0) + r_s S + r_e k_0 j - r_e \frac{k_0}{2T} j^2 + \alpha_1 Age + \alpha_2 Age^2 + \mu$$

This result held in both the men's equation and the women's equation, where the collinearity between the age and experience variables was less severe.

III. An Expanded Wage Model

Thus far the wage for an individual employee is assumed to be fully explained on the basis of schooling and postschooling investments. Perhaps the major missing elements in our explanation of salary differentials concern the various personal and productivity characteristics that influence the employment situation. It is, of course, very difficult to measure an individual's productivity independently of the salary decision. Nevertheless, proxy variables may be employed to estimate the individual's productivity and opportunities for alternative employment.

Given the scientific and technical nature of the output of the particular divisions of the corporation chosen for study, a variable measuring the extent of the individual's publications may serve as a useful proxy for productivity. An individual's value to the organization, and certainly his or her ability to generate outside job offers, is related to the employee's professional reputation. Individuals obtain outside reputations in scientific fields primarily through published work. Consequently we added a variable indicating the extent to which the individual had made important scientific contributions through his writing.

A cardinal measure of the individual's publications is not necessarily indicative of the employee's professional reputation or worth to the corporation. Consequently, we used a dummy variable, which took on the value "1" if the individual had made significant contributions and "0" otherwise. Several methods of coding the dummy variable were tried, based on the number of publications. All were successful, but the most satisfactory method combined a quantity and quality measure. The variable took on the value "1" if the individual had fifteen or more publications or, where publications were fewer than fifteen, a group of professionals in the individual's field judged that the publications were outstanding and so well known in the literature that he or she had achieved a distinguished outside reputation.

We also utilized a dummy variable indicating whether or not the individual had received the Ph.D. degree. Consider two individuals who spend four years each in graduate school. The first manages to obtain only a master's degree during the period while the second completes a dissertation and receives a doctorate. We suggest that the second individual is likely to be a more productive worker and will command a higher salary.

Another variable was added to indicate the individual's marital status. The variable was coded "1" if the individual was married and "0" if the individual was single. It has been argued by William Bowen and T. Aldrich Finegan that employers may regard an individual's marital status as a proxy for personal traits relating to success in the work environment. Family and marital responsibilities may make married men more stable than single men and, therefore, married men get the better-paying jobs. Also the employer may rationalize the payment of higher wages to married men because their financial requirements are greater. On the other hand, employers may expect married women to have higher absence and turnover rates, that is, they are "less stable" employees. For example, it is generally

presumed by employers that a working wife will quit her job when the husband's job changes so as to require the family to move.[7]

The actual number of years of schooling and on-the-job training may not capture adequately the quality of the investment made. The area in which an individual has undertaken undergraduate and graduate study may have an important bearing on salary prospects. For example, individuals with training in mathematics, statistics, and operations research would find many high paying opportunities available with universities, government agencies, or corporations. Other individuals trained in English literature might find only relatively low paying jobs available at universities or, perhaps, with publishing firms or similar employers. Thus, people with "critical" skills have a much higher opportunity cost and need larger salaries to be attracted to and retained by the organization.

In coding the employees with respect to the critical nature of their areas of study, we used a study, prepared for the American Economic Association, of median salaries for Ph.D.'s in alternative fields. The highest of these were regarded as critical areas of study. For the purposes of the present study, the fields of mathematics, statistics, and several of the physical sciences were coded as critical. A dummy variable, D_c, was employed which took on the value unity if the individual's area of study was a critical one, and zero otherwise.

Finally, each individual's absence rate, was added as an independent variable. Absence might have two effects on salaries. First, annual salaries for workers with a history of absence might be lowered to reflect the expectation that such employees would work fewer days per year. Second, absence may be related to looser work habits and may thus serve as a proxy for worker productivity. Workers who are frequently absent may be less productive even when they are on the job. The variable used in the regression was defined as simply the number of days absent per year other than vacation time, averaged over the preceding four years (or over the number of years employed if less than four).

As an expanded wage model, we therefore add to equation (3), the personal and productivity factors noted above. The broadened equation was

$$(5)\quad ln\,(W_t) = a_0 + a_1S + a_2j + a_3j^2 + a_4D_{PHD} + a_5D_{PUB} + a_6D_M + a_7D_c + a_8A + \eta$$

where D_{PHD}, D_{PUB}, D_M, and D_c stand for the Ph.D., publications, marital status, and critical area-of-study dummies, respectively, and A stands for the absence rate. Estimates of equation (5) for the years 1966, 1969, 1970, and 1971 are shown in Table 3.

Table 3 indicates that the additional "productivity" proxy variables do improve our ability to predict salary levels for men and women in the organization.[8] In most cases more than three-fourths of the variance in men's salary levels is explained by the regressions, while over 80 percent of the variance in women's salaries is explained. The coefficient of the dummy variable associated with publica-

[7] In fact, as will be shown below, the absence rate was much higher for women than for men, although marital status itself was not related to absence. A previous study of the organization was unable to isolate differences in turnover among men and women. It is for this reason we suggested earlier that the sample was relatively homogeneous with respect to the employee's attachment to the job.

[8] Utilizing an F-test similar to that described in fn. 6, we verified that the addition of the productivity variables was highly significant in all years. Nevertheless, the increase in R^2 over the narrow form of the model is only moderate, indicating the robustness of the cruder model.

TABLE 3—ESTIMATES OF EQUATION (5)[a]

Sample Group	Constant Term	Coefficient of S	j	j^2	D_{PHD}	D_{PUB}	D_M	D_c	A	R^2	F-Value	Standard Error of Estimate
Men 1971	9.008	+0.036	+0.044	−0.001	+0.118	+0.147	+0.145	+0.066	−0.005	0.68	31.83	0.161
		(2.77)	(7.28)	(5.18)	(2.67)	(2.91)	(2.88)	(1.99)	(1.09)		(8,120)	
Women 1971	8.719	+0.044	+0.061	−0.001	+0.047	+0.355	−0.013	+0.216	−0.015	0.83	49.83	0.155
		(3.88)	(7.90)	(4.65)	(0.81)	(3.54)	(0.39)	(5.36)	(4.62)		(8,84)	
Men 1970	8.790	+0.042	+0.051	−0.001	+0.121	+0.112	+0.183	+0.078	−0.001	0.76	53.88	0.156
		(3.41)	(10.09)	(7.10)	(2.95)	(2.40)	(4.18)	(2.50)	(1.24)		(8,133)	
Women 1970	8.692	+0.046	+0.059	−0.001	+0.044	+0.337	−0.010	+0.202	−0.015	0.84	61.67	0.143
		(4.63)	(9.12)	(5.15)	(0.82)	(3.64)	(0.28)	(5.61)	(5.26)		(8,93)	
Men 1969	8.767	+0.041	+0.048	−0.001	+0.141	+0.075	+0.162	+0.084	−0.004	0.78	65.33	0.148
		(3.79)	(11.05)	(7.37)	(3.95)	(1.74)	(4.30)	(3.04)	(1.08)		(8,150)	
Women 1969	8.692	+0.043	−0.054	−0.001	+0.078	+0.318	−0.012	+0.182	−0.014	0.82	58.43	0.149
		(4.31)	(10.04)	(6.03)	(1.46)	(3.31)	(0.40)	(5.00)	(4.78)		(8,104)	
Men 1966	8.828	+0.033	+0.049	−0.001	+0.105	+0.100	+0.034	+0.109	−0.048	0.77	29.30	0.149
		(2.35)	(7.35)	(4.97)	(2.31)	(1.77)	(0.51)	(2.72)	(1.68)		(8,70)	
Women 1966	8.361	+0.059	+0.074	−0.002	−0.015	+0.322	−0.027	+0.138	−0.013	0.83	34.35	0.160
		(4.69)	(8.70)	(6.11)	(0.18)	(2.43)	(0.61)	(2.83)	(3.51)		(8,55)	

[a] The numbers underneath the coefficients are the related t-statistics.
Note: S, j, are defined in Table 1; D_{PHD}=Ph.D.; D_{PUB}=publications; D_M=marital status; D_c=critical area-of-study; and A stands for the absence rate.

tions always had the correct sign and usually was highly significant. Similarly, the receipt of the Ph.D. and study in a critical area were typically associated with higher salary levels. Of course, the coefficients of S and j shrink and lose their former interpretation because the Ph.D. and publications variables represent unmeasured aspects of investment in S and j, respectively, and are correlated with them. Finally, the marital status dummies took on their expected signs, although the coefficients for women were never significant. The coefficient of the absence rate was typically negative and especially significant for women.[9]

[9] It should be noted that we were unsuccessful in finding a relationship between several other variables and the structure of salaries. We tried, for example, to measure the quality of education by obtaining the quality rating of the institution at which the education was received, and also by looking at whether or not the individual received academic honors. We could find no relationship between either variable and the structure of salaries. Our procedure was to add explanatory variables to the salary structure equations until further additions failed to reduce the standard error of the estimate adjusted for degrees of freedom.

As in the narrow form of the postschooling investment model, we were able to reject the hypothesis that the structural salary relationships estimated for men and women were the same. The null hypothesis that the two equations had the same slopes and constant terms was rejected at the 1 percent level in each year on the basis of the standard Chow test. While many of the slope coefficients were not significantly different, differences in the overall structure were caused by significant differences in the constant terms and in the coefficients of most of the dummy variables. As noted above, the similarity of the schooling and experience coefficients (compared with the narrow model) should not be taken to indicate that the rates of return for men and women are the same.

IV. Adjusted Salary Differentials or Discrimination Measures

Gross differences in average salaries between men and women (i.e., $\overline{W}_m - \overline{W}_w$) do not indicate the presence of discrimina-

tion. If women, on average, have less education and job experience than men, we should anticipate that their average salaries would be lower. Discrimination can be said to exist only if women with the same characteristics as men tend to receive lower salaries.

The basic technique involved in estimating salary differentials adjusted for differences in personal and productivity characteristics is to compare the average wages that would be received by the two groups if they were paid according to the same salary structure. Two different types of assumptions are possible. We could assume that in the absence of discrimination, both men and women would face the salary structure applicable for men, or that both sexes would face the salary structure applicable for women.

We define $\tilde{W}_w$ as the average salary women would receive if they were paid by the salary structure applicable for men. We may estimate $\tilde{W}_w$ by

$$(6) \qquad \tilde{W}_w = f_m(\bar{S}_w, \bar{j}_w, \bar{\phi}_w)$$

where f_m stands for the pay structure relationship estimated for men. It will be noted that equation (6) evaluates the function at the mean values of the schooling, experience, and ϕ_w, the vector of "productivity" variables applicable to the women in the sample. The gross difference in average salaries between men and women $(\overline{W}_m - \overline{W}_w)$ may be decomposed into two components: 1) the difference in salaries due to different characteristics, which may be estimated as $\overline{W}_m - \tilde{W}_w$, and 2) the differential remaining after adjustment for differing characteristics, i.e., $\tilde{W}_w - \overline{W}_w$, which may be interpreted as a measure of sex discrimination.[10]

Alternatively, we define $\tilde{W}_m$ as the average salary men would receive if they were paid according to the women's pay structure. Then $\tilde{W}_m$ may be estimated by

$$(7) \qquad \tilde{W}_m = f_w(\bar{S}_m, \bar{j}_m, \bar{\phi}_m)$$

where f_w, the pay structure function estimated for women, is evaluated at the mean values of the characteristics variables. As before, the gross difference in wages can be decomposed into that component due to differing characteristics, $\tilde{W}_m - \overline{W}_w$, and the adjusted difference remaining, $\overline{W}_m - \tilde{W}_m$.

Table 4 presents the average salaries for men and women for the period studied and a decomposition of the gross differentials. The top part of the table is based on the regressions estimating the narrow form of the postschooling investment model (equation (3)), the bottom part is based on the expanded model of equation (5). Women's salaries have consistently averaged only about two-thirds of the salaries for men. The gross difference in salaries, \$7,000 in 1971, should not be attributed to discrimination, however. Even if women were paid according to the same salary relationship existing for men, they would still earn only about 80 percent of the salary of men.

Large salary differentials would remain even if the same pay structure prevailed for men and women. The reasons are indicated in Table 5, which shows by sex the mean values for the independent variables of the wage-structure equation. The women in the sample are less endowed with the characteristics preferred by the organization. Women had less education, related work experience, and seniority than the men and more unrelated experience.[11] In addition, a smaller

[10] This approach is similar to the one employed by Ronald Oaxaca.

[11] Of course, culture, tradition, and even overt discrimination may explain why women do not have the same training and experience as men. Women may have been discouraged from graduate training and from particular fields of study such as mathematics and physics. They may, by choice, have devoted a number of years of their working lives to childbearing and household activities.

TABLE 4—ANALYSIS OF SALARY DIFFERENTIALS
NARROW POSTSCHOOLING INVESTMENT MODEL: EQUATION (3); EXPANDED MODEL: EQUATION (5)

	Equation	1971 Men A	1971 Women B	1971 B/A in Percent	1970 Men A	1970 Women B	1970 B/A in Percent	1969 Men A	1969 Women B	1969 B/A in Percent	1966 Men A	1966 Women B	1966 B/A in Percent
Average Salary		$20,273	$13,273	65.47	$18,030	$12,043	66.79	$16,357	$10,941	66.89	$14,608	$9,651	66.07
Salary if men and women were paid according to men's pay structure	(3)	20,273	16,447	81.13	18,030	14,380	79.76	16,357	13,088	80.01	14,608	12,160	83.24
	(5)	20,273	15,719	77.54	18,030	13,609	75.48	16,357	12,456	76.15	14,608	11,856	81.16
Salary if men and women were paid according to women's pay structure	(3)	15,121	13,273	87.78	13,486	12,043	89.30	12,099	10,941	90.43	10,362	9,651	93.14
	(5)	15,749	13,273	84.28	14,208	12,043	84.76	12,805	10,941	85.44	10,859	9,651	88.88
Analysis of Gross Differences in Percent													
Gross Difference $\overline{W}_m-\overline{W}_w$			7,000	100.00		5,987	100.00		5,416	100.00		4,957	100.00
Difference due to different characteristics $\overline{W}_m-\hat{W}_w$	(3)		3,826	54.66		3,650	60.97		3,269	60.36		2,448	49.38
	(5)		4,554	65.06		4,421	73.84		3,901	72.03		2,752	55.52
Adjusted Difference Remaining $\hat{W}_w-\overline{W}_w$	(3)		3,174	45.34		2,337	39.03		2,147	39.64		2,509	50.62
	(5)		2,446	34.94		1,566	26.16		1,515	27.97		2,205	44.48

percentage of women studied in critical areas, gained the Ph.D., or had published significantly. Finally, women tended to have considerably higher absence rates than men. As the bottom rows of Table 4 show, the major part of the gross difference in wages can be attributed to the differential training and other job characteristics women bring to the employment situation. Estimates of discrimination or "adjusted salary differentials" tended to be somewhat higher when salary structures are based on the narrow postschooling investment model. This is because women were less endowed than men with qualities we have suggested may proxy for worker productivity.

It may be argued that our analysis of adjusted salary differentials is biased because some of the independent variables measure different things in men and women. The primary example is marriage, which employers may tend to associate with maturity and stability for men and divided responsibilities for women. It may be unreasonable to say that in the absence of discrimination married women would earn more than single ones—but this is

TABLE 5—AVERAGES OF VARIABLES

Means	1971 Men	1971 Women	1970 Men	1970 Women	1969 Men	1969 Women	1966 Men	1966 Women
Annual wage	$20,273.36	$13,272.93	$18,029.79	$12,043.33	$16,356.98	$10,940.64	$14,608.23	$9,651.41
Level code	9.38	6.58						
Age	40.83	38.91	39.19	37.54	37.76	36.59	38.84	37.52
Years of seniority and related experience	14.86	11.86	13.53	10.62	12.03	9.76	13.06	10.41
Years of nonrelated experience	1.20	1.34	1.11	1.47	1.03	1.42	1.06	1.63
Years of higher education	6.73	5.07	6.66	5.11	6.70	5.12	6.89	5.20
Percent of sample with Ph.D. degree	0.53	0.20	0.52	0.19	0.53	0.18	0.53	0.19
Percent of sample classified as having publications	0.12	0.03	0.12	0.03	0.11	0.03	0.15	0.03
Percent of sample having critical areas of study	0.26	0.23	0.25	0.22	0.26	0.20	0.25	0.23
Number of days absent (average of past 4 years)	3.69	6.42	3.86	6.54	3.94	6.43	5.03	7.13
Percent of sample married	0.91	0.62	0.89	0.62	0.89	0.61	0.92	0.61

precisely what is assumed when the equation for men is used together with the means of the women's characteristics variables.

To judge the effects of such "bias" we ran the same experiment omitting the marital status variable. The resulting estimates of discrimination against women were actually a bit larger than those presented in Table 4 for the expanded model. The inclusion of marital status reduces estimated discrimination. Since far fewer women were married, including marriage as a desirable characteristic tended to make the "justifiable" difference in average men's and women's salaries larger.

Table 4 indicates that, in percentage terms, the adjusted salary differential in the organization has decreased since 1966 although it rose in 1970, compared with the preceding two years. Our estimates of discrimination, while substantial, are somewhat lower than those of other investigators, who have studied mainly the less skilled occupations. Fuchs finds that even after adjusting for education, women's salaries tend to be about 60 percent of those for men, and indicates that such a percentage can be found in the Bible, the code of Hammurabi, and in data on the price of slaves in the antebellum South. Oaxaca, using a human capital approach and survey data collected in February 1967, found somewhat greater adjusted salary differentials than those reported here for 1966. These comparisons suggest that estimates of discrimination may be less for a relatively homogeneous sample such as utilized here.

V. Analysis of Regression Results Using a "Job Level" Variable

Starting in 1971, the corporation provided a thirteen-point ranking of job levels, designed to indicate the individual's level of responsibility. Since job levels had salary ranges associated with them, it was expected that the salary and level variables would be highly correlated. Indeed, the simple correlation between the two variables was about 0.8. Consequently, when level was added as an independent variable to equation (5), many of the t-values associated with the independent variables became insignificant. Nevertheless, it was possible to include some of the independent variables from (5) together with level to explain the structure of salaries in 1971. The equation estimated was

$$\text{(8)} \quad ln\,(W_t) = a_0 + a_1S + a_2j + a_3j^2 + a_4D_{PUB} + a_5D_c + a_6A + a_7L + \epsilon$$

where L stands for the workers job level. The estimates of equation (8) are shown in the top half of Table 6. As expected, the coefficients of determination improve considerably compared with those reported previously.

The bottom half of Table 6 presents an analysis of the related salary differentials, as was shown in Table 4. Essentially, all the adjusted salary difference, or "discrimination," seems to disappear. The gross salary differentials can now fully be explained by differences in worker characteristics, including job level. What has happened is that women, on average, have considerably lower job levels than the men. Given the job levels to which women are assigned, they tend to make about the same salaries as men.

The question immediately arises, however, why women are assigned to much lower job levels than men. Can this be explained by their lower levels of training and education, or does the organization discriminate in assigning men and women to particular job levels? This can be analyzed by the same techniques employed above. We can try to explain assignment to job levels on the basis of the same types of employee characteristics

TABLE 6—ESTIMATES OF EQUATION (8) WITH JOB-LEVEL VARIABLE

1971

Sample Group	Constant Term	Coefficient of S	j	j^2	D_{PUB}	D_c	A	L	R^2	F-Value	Standard Error of Estimate
Men	8.582	+0.017 (3.19)	+0.016 (4.50)	−0.0003 (3.54)	+0.127 (4.73)	+0.038 (2.16)	−0.005 (1.90)	+0.109 (18.34)	0.91	166.66 (7,121)	0.087
Women	8.563	+0.019 (2.78)	+0.028 (4.80)	−0.0005 (3.13)	+0.281 (4.39)	+0.147 (5.46)	−0.006 (2.73)	+0.081 (10.83)	0.93	155.78 (6,85)	0.101

Analysis of Related Salary Differentials

	Men A		Women B	B/A in Percent
Average Salary	\$20,273		\$13,273	65.47
Salary if men and women were paid according to relationship for men	20,273		13,419	66.19
Salary if men and women were paid according to relationship for women	17,385		13,273	76.35
Gross Difference $\overline{W}_m - \overline{W}_w$		7,000		100.00
Difference due to different characteristics $\overline{W}_m - \tilde{W}_w$		6,854		97.91
Adjusted difference remaining $\tilde{W}_w - \overline{W}_w$		146		2.09

Note: the term L stands for workers' job level; S, j are defined in Table 1; D_{PUB}, D_c, A are defined in Table 3.

previously used to explain salary differentials. The specific equation employed was

$$(9)\quad L = \alpha_0 + \alpha_1 S + \alpha_2 j + \alpha_3 j^2 + \alpha_4 D_{PHD} + \alpha_5 A + \alpha_6 D_{PROD} + \alpha_7 D_M + \alpha_8 D_c + v$$

where D_{PROD} was a dummy variable, which took on the value unity if an individual employee worked in the "production" division, where job level assignments tended to be relatively low. The results of this exercise are shown in the top half of Table 7. Education, experience, and other characteristics can be used to make reasonably good predictions of an employee's job level.[12]

The bottom half of Table 7 presents an analysis of the differences in levels. The analysis is the same as was carried out for salary differences. On average, the level for men employees was 2.8 level steps higher than for women. Almost one-half of this gross difference can be attributed to the different characteristics of the men and women in the sample. Women can be expected to be 1.31 level steps lower than men because of their lower level of education, experience, etc. The remaining half of the difference remains unexplained and may be attributed either to discrimination against women or to other productivity characteristics, which differ by sex and which we were unable to measure.

VI. Conclusion

This study has attempted to answer two questions. 1) Can the structure of salaries in an individual organization be explained

[12] Equation (9) was reestimated under an assumption that the relationship between absence rate and level was a simultaneous one. We assumed that the absence rate affects level, but that the level assignment also affects the absence rate, particularly for women. We hypothesized that women assigned to low job levels (inappropriate to their training) might be bored with their assignments and, therefore, be absent more. Thus, the absence rate, A, was hypothesized to depend upon age and level, L, in the following fashion:

$$A = \beta_0 + \beta_1 Age + \beta_2 Age^2 - \beta_3 L + \delta$$

When equation (9) was reestimated with a two-stage least squares technique, the coefficient estimates were essentially the same as those reported above. Moreover, the estimated coefficient, β_3, of the absence rate equation above was not significantly different from zero. We did find, however, that the age was important in explaining absence. The estimated coefficient β_2 was positive indicating that the older employees were more likely to be absent from work. The age variables explain only a small part of the variance in absence rates, however. Absence was found to be unrelated to such plausible explanatory variables as marital status and presence of children under six years of age for women.

Table 7—Estimates of Equation (9) Predicting Job Levels

1971

Sample Group	Constant Term	Coefficient of S	j	j^2	D_{PHD}	A	D_{PROD}	D_M	D_c	R^2	F-Value	Standard Error of Estimate
Men	4.025	+0.259 (2.52)	+0.255 (5.34)	−0.004 (3.68)	+0.668 (1.88)	−0.010 (2.69)	−0.533 (1.79)	+0.924 (2.35)	+0.233 (0.89)	0.53	17.06 (8,120)	1.271
Women	1.741	+0.349 (3.13)	+0.428 (5.82)	−0.008 (3.49)	+0.701 (1.31)	−0.115 (3.70)	−0.341 (0.95)	−0.012 (0.04)	+0.808 (2.12)	0.69	23.38 (8,84)	1.469

Analysis of Level Differentials

	Men A		Women B	B/A in Percent
Average Level	9.38		6.58	70.15
Level if men and women were assigned job levels according to relationship for men	9.38		8.07	86.03
Level if men and women were assigned job levels according to relationship for women	8.24		6.58	79.85
Gross Difference in Levels		2.80		100.00
Difference due to differing characteristics of men and women in sample (assuming men and women were assigned levels according to the relationship for men)		1.31		46.79
Difference remaining		1.49		53.21

Note: D_{PROD} stands for workers in production; S, j are defined in Table 1; D_{PHD}, D_M, D_c, A are defined in Table 3.

by such variables as education, experience, and productivity proxies? 2) Is there discrimination against women; that is, do women with job characteristics equal to those of men fail to get the same pay?

We have been quite successful in answering to the first question. Experience, education, and productivity variables are able to explain over three-quarters of the variance in salaries for both men and women over several years. The answer to the second question depends upon how narrowly or broadly one perceives the problem. If the question is posed, "Do men and women in equal job levels, with the same characteristics, get equal pay?" the answer is "Yes." If, on the other hand, the question is posed, "Do men and women, with equal characteristics, get equal pay?" the answer is "No." The reason for the difference in answers is that women with the same training, experience, etc. as men tend to be assigned to lower job levels.

It is, of course, possible that such a pattern of job assignment does not reflect discrimination. Other explanations may be preferences of men and women for more or less responsibility or biases resulting from our culture and tradition which may affect productivity characteristics of men and women other than those that we were able to measure and include in the study. It seems highly improbable, however, that such missing variables would fully account for the pattern of job level assignments or for the adjusted salary differentials we have isolated. We suggest that it is difficult for a discriminating organization to give male and female employees the same titles and pay them different amounts. It is far easier to assign women to lower job levels and then set up a pay structure by level that is the same for both sexes. Thus, our analysis of salary differentials including job levels should not be interpreted as indicating an absence of discrimination. The assignments to job levels can most plausibly be interpreted as the mechanism by which the discrimination takes place.

REFERENCES

G. S. Becker, *Human Capital and the Personal Distribution of Income: An Analytic Approach,* Woytinsky Lecture No. 1, Univ. Michigan 1967.

——— and B. R. Chiswick, "Education and the Distribution of Earnings," *Amer. Econ. Rev. Proc.*, May 1966, *56*, 350–69.

W. G. Bowen and T. A. Finegan, *The Economics of Labor Force Participation*, Princeton 1969.

G. C. Chow, "Tests of Equality Between Sets of Coefficients in Two Linear Regressions," *Econometrica*, July 1960, *28*, 591–605.

V. R. Fuchs, "Male-Female Differentials in Hourly Earnings," Nat. Bur. Econ. Res. 1970, unpublished.

T. Johnson, "Returns from Investment in Human Capital," *Amer. Econ. Rev.*, Sept. 1970, *60*, 546–60.

J. Mincer, "The Distribution of Labor Incomes: A Survey with Special Reference to the Human Capital Approach," *J. Econ. Lit.*, Mar. 1970, *8*, 1–26.

R. L. Oaxaca, "Male-Female Wage Differentials in Urban Labor Markets," Princeton Univ. Industrial Relations Section Working pap. no. 23, 1971.

N. A. Tolles et. al., Committee on the National Science Foundation Report on the Economics Profession, "The Structure of Economists' Employment and Salaries, 1964," *Amer. Econ. Rev.*, Dec. 1965, supp., *55*.

Nontraded Goods, Factor Market Distortions, and the Gains from Trade

By RAVEENDRA N. BATRA*

The purpose of this paper is to explore one by one the implications of a change in the rate of tariff and the terms of trade for community welfare in an economy where some goods are traded and some nontraded and where factor markets are distorted by the presence of invariant interindustry wage differentials. Despite the claim to the contrary by Murray Kemp,[1] we find that an increase in the rate of tariff may actually lead to a rise in welfare even when distortions are absent. Another major result of this note is that, unlike the results derived by Jagdish Bhagwati and V. K. Ramaswami, free trade alone may turn out to be the second best policy in the presence of distortions, provided the higher wage rate is paid by the import-competing industry or the wage rates differ among all industries. Furthermore, unlike the traditional two-good models, factor proportions may play a significant role in determining the fate of several standard theorems concerning the gains from trade.

I. Assumptions and the Model

Unless otherwise specified, we assume that there is a small country which produces three goods, X_1, X_2, and X_3, with the help of given supplies of two factors of production, capital (K) and labor (L). In its trade relations with the rest of the world it takes the international price ratio as given, exports X_2 and imports X_1. The third good, X_3, is not traded. There is perfect competition in product markets, and production functions are homogeneous of the first degree. Factor markets are characterized by a) perfect internal mobility of factors, and perfect factor-price flexibility, and b) the reward of capital is the same in all three sectors, but the wage rate differs among all three industries *in spite of the factor mobility*. There is full employment of factors, which is ensured by the presence of factor-price flexibility. Foreign trade is stable and inferior goods are absent. In addition all goods are gross substitutes.

Let U denote the total utility which a community derives from the consumption of the three goods, with their demand denoted by D_i ($i=1, 2, 3$). Then

(1) $$U = U(D_1, D_2, D_3)$$

(2) $$D_1 = X_1 + E_1$$

(3) $$D_2 = X_2 - E_2$$

(4) $$D_3 = X_3$$

where E_1 equals the import of the first commodity and E_2 equals the export of the second commodity. The balance of trade equilibrium requires that

(5) $$P_1E_1 = P_2E_2$$

where P_j is the given foreign price of the jth traded good ($j=1, 2$). The three linearly homogeneous production functions are given by

(6) $$Xi = F_i(K_i, L_i) = L_if_i(k_i) \quad (i = 1, 2, 3)$$

where K_i and L_i are the capital and labor inputs and $k_i=K_i/L_i$ is the capital-labor ratio in the ith commodity. Let F_{Ki} and F_{Li} be the marginal productivity of capital and labor, respectively, in the ith commodity. Then $F_{Ki}=f_i'$ and $F_{Li}=f_i-k_if_i'$. The price of each factor of production equals the value of its marginal product. With the reward of capital assumed to be the same in all industries,

(7) $$r = P_1f_1' = P_2f_2' = P_3f_3'$$

* Professor of economics, Southern Methodist University. I wish to express my gratitude to V. S. Rao for useful discussions and to the editor and an unknown referee for suggesting several refinements. This research was supported by a grant from Canada Council.

[1] See Kemp, ch. 12, note 13, p. 267.

where P_3 is the price of the nontraded good. However, in the presence of the wage differentials in all industries

$$(8)\qquad \frac{P_1(f_1 - k_1 f_1')}{\alpha_1} = \frac{P_2(f_2 - k_2 f_2')}{\alpha_2} = \frac{P_3(f_3 - k_3 f_3')}{\alpha_3}$$

where $\alpha_i \geq 1$ $(i=1, 2, 3)$ is a constant. Under full employment,

$$(9)\qquad L_1 + L_2 + L_3 = L$$

and

$$(10)\qquad K_1+K_2+K_3=L_1k_1+L_2k_2+L_3k_3=K$$

This completes the specification of our model which allows for the presence of the factor-price differential as well as a nontraded good.

The Relationships Among X_1, X_2, and X_3

For a small country P_1, P_2, and hence P_3 are given, so that k_i is constant.[2] Differentiating X_1 and X_2 from (6) with respect to X_3, we obtain

$$\frac{dX_i}{dX_3} = f_i \frac{dL_i}{dX_3} \qquad (i = 1, 2)$$

Substituting this in the differentiation of the full employment equations (9) and (10) then yields

$$(11)\qquad \frac{dX_1}{dX_3} = \frac{f_1(k_3 - k_2)}{f_3(k_2 - k_1)}$$

$$(12)\qquad \frac{dX_2}{dX_3} = \frac{f_2(k_1 - k_3)}{f_3(k_2 - k_1)}$$

Equations (11) and (12) reveal the importance of factor proportions in the three industries in the determination of the relationships among X_1, X_2, and X_3. For expository purposes, we also need the relationship between the production-possibility surface and the price plane in the presence of the inter-industry wage differentials.

[2] Ryutaro Komiya has shown that if P_1 and P_2 are constant, P_3 also remains constant.

Totally differentiating X_1 and X_2 in (6) and dividing through by dX_2, we obtain

$$(13)\qquad \frac{dX_1}{dX_2} = \frac{F_{K1}dK_1 + F_{L1}dL_1}{F_{K2}dK_2 + F_{L2}dL_2}$$

From (9) and (10),

$$dL_1 + dL_2 + dL_3 = dK_1 + dK_2 + dK_3 = 0$$

Using these relations and (7) and (8), (13) can be written as

$$(14^*)\qquad \frac{dX_1}{dX_2} = -\frac{P_2}{P_1} \cdot \left[\frac{F_{K2}(dK_2+dK_3)+(\alpha_1/\alpha_2)F_{L2}(dL_2+dL_3)}{F_{K2}dK_2+F_{L2}dL_2}\right]$$

or

$$(14)\qquad \frac{dX_1}{dX_2} = -\frac{P_2}{P_1} \cdot \left[\beta + \frac{F_{K2}dK_3 + (\alpha_1/\alpha_2)F_{L2}dL_3}{F_{K2}dK_2 + F_{L2}dL_2}\right]$$

where

$$\beta = \frac{F_{K2}dK_2 + (\alpha_1/\alpha_2)F_{L2}dL_2}{F_{K2}dK_2 + F_{L2}dL_2} \gtreqless 1$$

according as $(\alpha_1/\alpha_2) \gtreqless 1$. This follows from the fact that dK_i and dL_i and hence the numerator and the denominator of β have the same signs.[3] In the same manner, we can write that

$$\frac{dX_3}{dX_2} = \frac{P_2}{P_3}\left[\frac{F_{K2}dK_3 + (\alpha_3/\alpha_2)F_{L2}dL_3}{F_{K2}dK_2 + F_{L2}dL_2}\right] = \frac{P_2}{P_3}\left[\frac{F_{K2}dK_3 + (\alpha_1/\alpha_2)F_{L2}dL_3}{(F_{K2}dK_2 + F_{L2}dL_2)\lambda}\right]$$

where

$$\frac{1}{\lambda} = \frac{F_{K2}dK_3 + (\alpha_3/\alpha_2)F_{L2}dL_3}{F_{K2}dK_3 + (\alpha_1/\alpha_2)F_{L2}dL_3} \gtreqless 1$$

[3] This can be readily seen by comparing the production points along the contract curve in the Edgeworth-Bowley production box diagram in the two-good, two-factor case and in the Melvin type box diagram in the three-good, two-factor case. The reader will find that the rise in the output of the ith good is accompanied by the rise in both K_i and L_i, and conversely.

if $(\alpha_3/\alpha_2)\gtreqless(\alpha_1/\alpha_2)$, or $\lambda 1, \gtreqless 1$ according as $(\alpha_1/\alpha_3)\gtreqless 1$. It may now be seen that the second term within the brackets of (14) equals $(\lambda P_3/P_2)(dX_3/dX_2)$. Hence

$$(15^*) \quad \frac{dX_1}{dX_2} = -\frac{P_2}{P_1}\left[\beta + \frac{\lambda P_3}{P_2}\cdot\frac{dX_3}{dX_2}\right]$$

or

$$(15) \quad P_1dX_1 + \beta P_2dX_2 + \lambda P_3dX_3 = 0$$

Equation (15) shows that the price plane is not tangential to the production-possibility surface. In the absence of the wage differential, however, $\alpha_i=\beta=\lambda=1$, so that (15) reduces to $P_1dX_1+P_2dX_2+P_3dX_3=0$.

II. Welfare and the Change in the Terms of Trade

In this section we examine the Batra-Pattanaik theorem (derived in the absence of the nontraded goods) that an improvement in the terms of trade may result in a loss of welfare, and vice versa, if the wage differential is paid by the importable good. In our model, X_2 is the exportable good and X_1 the importable good. Therefore an exogenous improvement (deterioration) in the terms of trade can result from a rise (fall) in P_2 alone, or a fall (rise) in P_1 alone. Let us consider the implications for welfare of a slight change in P_2 only, keeping P_1 constant. Differentiating the social utility function given by (1) totally with respect to P_2, and dividing through by U_1, we get

$$(16) \quad \frac{1}{U_1}\,\frac{dU}{dP_2} = \frac{dD_1}{dP_2} + \frac{U_2}{U_1}\cdot\frac{dD_2}{dP_2} + \frac{U_3}{U_1}\cdot\frac{dD_3}{dP_2}$$
$$= \frac{dD_1}{dP_2} + \frac{P_2}{P_1}\cdot\frac{dD_2}{dP_2} + \frac{P_3}{P_1}\cdot\frac{dD_3}{dP_2}$$

where $U_i=\partial U/\partial D_i>0$, $(U_2/U_1)=(P_2/P_1)$ and $(U_3/U_1)=(P_3/P_1)$. Differentiating (2)–(5) with respect to P_2, dividing through (15) by P_1dP_2, substituting dX_1/dP_2 in (16) and remembering that $(dD_3/dP_2)=(dX_3/dP_2)$, we obtain

$$(17) \quad \frac{1}{U_1}\,\frac{dU}{dP_2} = \frac{E_2}{P_1} + \frac{P_2}{P_1}\cdot\frac{dX_2}{dP_2}(1-\beta) + \frac{P_3}{P_1}\,\frac{dD_3}{dP_2}(1-\lambda)$$

This last equation furnishes the long sought after relationship between the terms of trade and welfare measured in terms of the first commodity. The following results may now be derived.

1. Suppose the wage differential exists, but the nontraded good does not. Following Batra and Pattanaik, it is assumed that $dX_2/dP_2>0$ in spite of the wage differential. Then if $\beta>1$, it can be easily seen that an improvement (deterioration) in the terms of trade may lead to a loss (rise) in welfare. Now $\beta>1$ implies that $\alpha_1>\alpha_2$, which from (8) means that the differential (or the higher wage) is paid by importable goods. This is a simple proof of the Batra-Pattanaik result and is obtained as a special case of the general equation (17).

2. Suppose that the nontraded good exists. We have already established that $\beta\gtreqless 1$, if $(\alpha_1/\alpha_2)\gtreqless 1$ and that $\lambda\gtreqless 1$, if $(\alpha_1/\alpha_3)\gtreqless 1$. Let us first consider the case where the differential is paid by the nontraded good. Here $1=\alpha_1=\alpha_2<\alpha_3$, so that $\beta=1$ and $\lambda<1$. Under our assumption that all commodities are gross substitutes, $(dD_3/dP_2)>0$. Hence from (17), we can see that $(1/U_1)(dU/dP_2)>0$. *Thus we conclude that if the wage differential is paid by the nontraded good, and if the commodities are gross substitutes, then an improvement (deterioration) in the terms of trade results in an improvement (deterioration) in welfare.* This is a straightforward generalization of the theorem by Anne Krueger and Hugo Sonnenschein.[4] What is more interest-

[4] That the wage differential is paid by the nontraded good sector is not a mere theoretical possibility. There is a logical explanation for why the contrary may be untrue. Since the traded goods face strong competition from foreign producers which the nontraded goods do not, the employers in the nontraded goods may be more prone to accept union demands for higher wages than the employers in the trade goods. Recent restrictions introduced in the United States by President Nixon on the spiralling wage gains by unions in the construction industry, which is a very good example of a nontraded good industry, would lend credence to our

ing is that the change in welfare is even greater than what would result in the absence of the wage differential.

3. Of greater interest is the case where the differential is paid by the import competing industry so that $1=\alpha_2=\alpha_3<\alpha_1$. Here the results turn out to depend crucially on factor-proportions in all three industries. For by using (12), (17) can be written as

$$(18)\quad \frac{1}{U_1}\cdot\frac{dU}{dP_2}=\frac{E_2}{P_1}+\frac{P_3(1-\lambda)}{P_1}\frac{dD_3}{dP_2}\cdot\left[1+\frac{P_2(1-\beta)}{P_3(1-\lambda)}\cdot\frac{f_2(k_1-k_3)}{f_3(k_2-k_1)}\right]$$

With the wage differential paid by the importable good only, λ and β are both greater than unity, so that $(1-\lambda)<0$ and $(1-\beta)/(1-\lambda)>0$. The effect of a change in P_2 on welfare then depends on the sign of the terms within the brackets, and hence on the factor proportions in the three industries. A sufficient condition for the terms of trade to have a positive relationship with social welfare is that the square bracketed term is zero or negative, that is

$$(19)\qquad \frac{P_3(1-\lambda)}{P_2(1-\beta)}\leq\frac{f_2(k_3-k_1)}{f_3(k_2-k_1)}$$

Since the left-hand side of (19) is positive, a necessary condition for (19) to be satisfied is that its right-hand side be positive, which in turn requires that either $k_1\gtrless k_3\gtrless k_2$ or $k_1\gtrless k_2\gtrless k_3$. However, condition (19) is violated if k_1 lies between k_2 and k_3, so that with $\lambda>1$, an improvement in the terms of trade may then lead to a loss of welfare, and conversely. This is a variant of the Batra-Pattanaik theorem and it highlights the role of factor proportions of the three industries in determining the implications of a change in the terms of trade for welfare, when the differential is paid by the importable good and all goods are gross substitutes.

One may also wish to analyze the effects of the wage differentials prevailing in all the industries. In this case, the results derived above may be reinforced or attenuated. It is of some interest to note here that the effects of the wage differentials among all three industries may cancel out each other and leave the standard theorem unscathed. The necessary condition for this "cancelling effect" is again given by condition (19) with equality.

III. Higher Versus Lower Tariffs

This section is concerned with the implications of a rise in the rate of tariff on welfare. Assume that a small tariff already exists on the imports of the first commodity, so that $P_1^*=P_1(1+t)$, where P_1^* is the domestic, tariff-inclusive price for the importable commodity and t is the rate of nonprohibitive tariff. With P_1 constant, $dP_1^*/dt=P_1>0$. Differentiating (1)–(5) with respect to t and remembering that here $U_2/U_1=P_2/P_1^*$ and $U_3/U_1=P_3/P_1^*$, we obtain

$$(20)\quad \frac{1}{U_1}\frac{dU}{dt}=\left[\frac{t}{(1+t)}\right]\frac{dE_1}{dt}+\frac{dX_1}{dt}+\frac{P_2}{P_1^*}\frac{dX_2}{dt}+\frac{P_3}{P_1^*}\frac{dD_3}{dt}$$

With the imposition of the tariff, P_1 is replaced by P_1^* in (15), so that

$$(21)\quad P_1^*dX_1+\beta P_2dX_2+\lambda P_3dX_3=0$$

Dividing through (21) by P_1^*dt, substituting for dX_1/dt in (20) and with some manipulation, we obtain

$$(22)\quad \frac{1}{U_1}\cdot\frac{dU}{dt}=\frac{1}{1+t}\left[P_1t\frac{dE_1}{dP_1^*}+P_2\frac{dX_2}{dP_1^*}(1-\beta)+P_3\frac{dD_3}{dP_1^*}(1-\lambda)\right]$$

Equation (22), which indicates the effects of a change in the tariff rate on welfare, may

hypothesis. Similarly, in Canada the wage increases secured by unions in the housing industry have far outpaced the increases in other sectors, some of which do enter international trade. Thus, if there was no wage differential in the economy before, one will now be created in such a way that higher wages will be paid by the nontraded good sector.

now be utilized to derive the following results:

1. To begin with, consider these effects in the absence of the wage differential, so that with $\beta=\lambda=1$, (22) reduces to

$$\frac{1}{U_1}\frac{dU}{dt}=\frac{P_1 t}{1+t}\cdot\frac{dE_1}{dP_1^*} \tag{23}$$

In the absence of nontraded goods, $dE_1/dP_1^*<0$, that is, a rise in the domestic price of the importable good resulting from the rise in tariff lowers the demand for imports, so that $(1/U_1)(dU/dt)<0$. In other words, a rise in the tariff rate leads to a decline in welfare. This is the standard theorem derived by Kemp. However, when a nontraded good is introduced, the sign of dE_1/dP_1^* becomes ambiguous, because now the repercussions of a change in the domestic price of the importable good on the price of X_3 have also to be considered. Even if all goods are assumed to be gross substitutes, Komiya has shown that the sign of (dE_1/dP_1^*) may be uncertain, simply because of the ambiguity in the sign of (dX_1/dP_1^*), which is possible when a third, nontraded good is incorporated into the model. Thus we conclude that the effects of a rise in the tariff rate on social welfare are uncertain in the presence of the nontraded good. *What is of great interest is that a higher tariff may lead to higher welfare than a lower tariff in the presence of the nontraded good, in spite of the absence of the wage differential, so that the standard theorem by Kemp may not hold.*

2. When wage differentials are also introduced, (22) shows that the sign of $(1/U_1)(dU/dt)$ becomes ambiguous even if (i) $(dE_1/dP_1^*)<0$, (ii) $(dX_2/dP_1^*)<0$ and (iii) $dD_3/dP_1^*>0$, so that all goods are gross substitutes. The increase in the tariff rate may or may not cause a decline in welfare. Here again by following the procedure established in the previous section, one can demonstrate the significance of factor proportions in the three industries for the change in welfare when the higher wage rate is paid by the import-competing industry. The interested reader can derive his own conclusions with the help of condition (19).

If the wage differentials exist among all three industries, the results derived above may be strengthened or weakened. Again it may be noted that the effects of the wage differential may cancel out and the last two terms in the square brackets of (22) may add to zero. But the standard theorem by Kemp may still be invalid because of the possibility of the positive sign of (dE_1/dP_1^*).

IV. Free Trade Versus No Trade and Optimum Policy

In the last section we have shown that there is no monotonic relationship between the tariff rate and the level of social welfare when the interindustry wage differential is present even if (dE_1/dP_1^*) has the "normal" negative sign. Since free trade represents the absence of the tariff and no trade may be considered to be a situation with prohibitive tariff, the results derived in the previous section are applicable to the questions under consideration in the present section. In the absence of the monotonic relationship between the rate of tariff and social welfare, we conclude that in the presence of the interindustry wage differential, free trade may or may not be superior to no trade.[5]

In other words, the Bhagwati-Ramaswami theorem continues to hold when a nontraded good is introduced in their model incorporating factor market imperfections, although their condition that the wage differential should be paid by the importable good ceases to be the necessary condition,[6] for the sign of

[5] It is worth pointing out here that this procedure does not apply to the case where the wage differential does not exist. As shown above, (22) now reduces to $(1/U_1)(dU/dt)=(P_1t/1+t)(dE_1/dP_1^*)$, and the ambiguity in the sign of (dE/dP_1^*) in the presence of the nontraded good may lead one to the faulty conclusion that free trade may be inferior to no trade. In the subsequent sections free trade is shown to be the optimal policy in the absence of the wage differential regardless of the presence of nontraded goods. A simple proof of this result actually follows from (23) if we assume that there is no tariff in the initial situation, so that $t=0$ and with $\beta=\lambda=1$, $(1/U_1)(dU/dt)=0$. In the presence of the wage differential, however, it is easily seen that this is not true.

[6] The earlier trade literature on wage differentials has

$(1/U_1)(dU/dt)$ turns out to be indeterminate if the differential is paid by only the nontraded good. This brings us directly to the question of optimal policy in the presence of the wage differential and the nontraded good. Here it is fruitful to draw a distinction between the optimal policy in the presence of a stable interindustry factor-price differential, which we shall call the second best policy, and the optimal policy that serves to eliminate the factor-price differential to the producers, so that the *optimum optimorum* is achieved. This latter policy will be referred to as the first best policy. Assume that initially there is free trade. Differentiating (1)–(5) totally and following essentially the same procedure as that used in Section II, we obtain

$$\frac{dU}{U_1} = \frac{P_2}{P_1} dX_2(1-\beta) + \frac{P_3}{P_1} dX_3(1-\lambda) \tag{24}$$

The following results are then immediate:

1. The necessary condition for an interior maximum is that $(dU/U_1)=0$. If the wage differential does not exist so that $\beta=\lambda=1$, this condition is automatically satisfied. In other words, free trade is the first best policy in the absence of the distortionary wage differential even when a nontraded good is introduced.

2. If the nontraded good does not exist, then $(dU/U_1)=P_2dX_2(1-\beta)/P_1$. Here β reflects the divergence between the marginal rate of transformation and the international-price ratio, because from (15*) $(dX_1/dX_2) = -(\beta P_2/P_1)$ when the nontraded goods are absent. The optimal policy in the presence of the wage differential requires a change in X_2 such that the divergence between the marginal rate of transformation and the given international-price ratio is eliminated, so that the value of β to the producers comes to equal unity. The second best policy in other words requires the imposition of a production-tax-cum-subsidy on second commodity. This result is also due to Bhagwati and Ramaswami. The first best policy, of course, consists in the elimination of the effective wage differential to the producers by means of the factor tax-cum-subsidy policy, such that (α_1/α_2) and β are equated to one. This is the proof of the first best policy prescribed by Bhagwati and Ramaswami.

3. Suppose the nontraded good exists. Then if the wage differential is paid either by the nontraded or the exportable good, so that only β or λ differs from unity, the second best policy requires respectively the imposition of the production tax-cum-subsidy on the second commodity, or a *consumption subsidy* on the third commodity such that the divergence to the producers between $-(dX_1/dX_2)$ and

$$[(P_2/P_1)(1 + (P_3/P_2)(dX_3/dX_2)]$$

is eliminated, that is to say, the effective value of β or λ, as the case may be, is equalized to unity. Note that when the differential is paid by the nontraded good, the second best policy calls for the grant of a consumption subsidy to the domestic good, a policy that does not apply to the case of traded goods. Since the distortion is in the domestic production sector of the economy, the thrust of the cure should also lie in re-

tended to make an illusory distinction between a wage differential which leaves the pattern of trade unchanged and one that causes a change in this pattern. With the former type of wage differential, the beneficent effects of free trade continue to exist, whereas with the latter type, they may not (see Everett Hagen, Bhagwati and Ramaswami, etc.). However, it appears that this is an unsatisfactory way of presenting the conclusions, for whether or not the differential causes a reversal in the earlier pattern of trade, there are always two possibilities, namely, 1) the differential is paid by the exportable good, and 2) it is paid by the importable good. Suppose in the absence of the wage differential, one commodity is imported, and suppose now that the differential is introduced which is unfavorable to this importable good. Obviously, the earlier pattern of trade cannot be reversed in this case. The conventional literature would then create the impression that the standard theorems will now be valid because the earlier pattern of trade is not reversed, which is not true. One does not have to bother about the reversal of the pattern of trade. In a two-good model, as stated above, there are only two possibilities: under one the traditional results hold, under the other, they may not. In this paper, this is how the conventional results and our own results have been presented.

moving the inequality between $-(dX_1/dX_2)$ and $(P_2/P_1)[1+(P_3/P_2)(dX_3/dX_2)]$. Just as in the two-good model, the second best optimum is achieved by changing the output of X_1 and X_2 in such a way that the difference between $-(dX_1/dX_2)$ and (P_2/P_1) is eliminated. If the consumption tax (or subsidy) were to be imposed on one of the traded goods, there will not be any change in outputs, for (P_2/P_1) which is exogenously determined will remain unchanged for the producers. However, in the case of the nontraded good, the domestic output of X_3 equals its domestic consumption. Therefore, if the wage differential is paid by X_3 producers, the second best optimum will be attained only by increasing its output to such an extent that to its producers λ comes to equal unity. However, in order to raise the output of X_3 without creating a discrepancy in its domestic demand and supply, a consumption subsidy to the nontraded good, rather than a production subsidy, is required. The output of X_3 will automatically rise when its demand increases, for in the three-good, two-factor model, as demonstrated by James Melvin and Kemp, any number of output configurations are compatible with a given set of commodity prices. The production possibility surface is completely described by straight lines and each line is a locus of possible output configurations for the three commodities corresponding to a given set of commodity prices. Hence when a consumption subsidy is given to the nontraded good, a new production point can be selected for the given set of prices in such a way that $P_1dX_1+P_2dX_2+P_3dX_3=0$. It may also be noted from (24) that when $\lambda<1$, so that the wage differential is paid by the nontraded good, a positive sign of $dD_3=dX_3$ will raise welfare.

On the other hand, if a production subsidy was granted to the nontraded good, the change in its output would not necessarily be matched by the change in its consumption, so that a disequilibrium situation could arise which will not be self correcting because the price of the nontraded good to the consumers remains constant under the small country assumption.

4. Finally, if the wage differential is paid by the importable good, then, given that the nontraded good exists, both β and λ exceed unity. An interesting but chance result emerges here. It is possible that $(dU/U_1)=0$, provided $P_2dX_2(1-\beta)+P_3dX_3(1-\lambda)=0$, or

$$\frac{P_3(1-\lambda)}{P_2(1-\beta)}=\frac{dX_2}{dX_3} \tag{25}$$

In this case free trade alone is the second best policy and the imposition of the production tax-cum-subsidy is not needed. Although this result is purely of academic interest, for only by coincidence will β and λ be such as to satisfy (25), yet the condition does suggest an interesting possibility. In any case if the difference between the left- and right-hand sides of (25) is small, the gain to be derived from the introduction of the production tax-cum-subsidy may not be more than negligible and may not be worth the administrative headaches and costs. A necessary condition for (25) to be satisfied is that (dX_2/dX_3) be positive, because both β and λ exceed unity when the differential is paid by X_1 producers. The chances of (25) being satisfied improve if the differential exists among all three industries so that $\alpha_1\neq\alpha_2\neq\alpha_3$; nor is it necessary anymore that (dX_2/dX_3) be positive, because it is possible that $\lambda\lessgtr 1$ when $\beta\gtrless 1$. The interesting point is that there exists a certain configuration of wage differentials which will satisfy (25). Nevertheless, in all the cases discussed above, the first best policy requires the use of factor tax-cum-subsidy so that the differential to the producers is eliminated.

An interesting policy device to attain the second best welfare suggests itself from this discussion. Suppose the level of the inter-industry factor-price differential itself depends on the government policy. For example, Harry Johnson, referring to Arnold Harberger, has suggested that the inter-industry factor-price differential will occur if the government imposes a corporation income tax (which is a tax on the use of capital in the corporate sector alone), that drives a wedge between the gross rewards of capital

in the corporate and the noncorporate sector. If the government must impose the corporation income tax in order to raise revenue, etc., it could be imposed in such a way as to satisfy condition (25). For a small country, the commodity prices and hence factor-proportions and f_2 and f_3 are given. All we then need to do is to select that level of tax on the use of capital, and hence the level of λ and β, such that $(1-\lambda)/(1-\beta)$ is equated to $P_2f_2(k_3-k_1)/P_3f_3(k_2-k_1)$, because from (12) $(dX_2/dX_3)=f_2(k_3-k_1)/f_3(k_2-k_1)$. Although λ and β represent the wage differential among the three industries under the assumption of similar reward of capital in all industries, the arguments are little modified if the rewards of capital differ among industries but the wage rates are similar.

REFERENCES

R. **Batra** and **P. K. Pattanaik**, "Domestic Distortions and the Gains from Trade," *Econ. J.*, Sept. 1970, *80*, 638–49.

J. **Bhagwati** and **V. K. Ramaswami**, "Domestic Distortions, Tariffs, and the Theory of Optimum Subsidy," *J. Polit. Econ.*, Feb. 1963, *71*, 44–50.

E. E. **Hagen**, "An Economic Justification of Protectionism," *Quart. J. Econ.*, Nov. 1958, *72*, 496–514.

A. C. **Harberger**, "The Incidence of the Corporation Income Tax," *J. Polit. Econ.*, Apr. 1962, *70*, 215–40.

H. G. **Johnson**, "Factor Market Distortions and the Shape of the Transformation Curve," *Econometrica*, July 1966, *34*, 686–98.

M. C. **Kemp**, *The Pure Theory of International Trade and Investment*, Englewood Cliffs 1969.

R. **Komiya**, "Non-Traded Goods and the Pure Theory of International Trade," *Int. Econ. Rev.*, June 1967, *8*, 132–52.

A. O. **Krueger** and **H. Sonnenschein**, "The Terms of Trade, the Gains from Trade and Price Divergence," *Int. Econ. Rev.*, Feb. 1967, *8*, 121–27.

J. R. **Melvin**, "Production and Trade with Two Factors and Three Goods," *Amer. Econ. Rev.*, Dec. 1968, *58*, 1249–68.

Optimal Effective Protection in General Equilibrium

By Francisco R. Casas*

In a well known article, Roy Ruffin has shown that for a small country importing a final as well as an intermediate product, the second best optimum effective tariff rate using the Barber-Johnson-Balassa measure is zero when the imported input is not produced domestically and is used in the import-competing sector only.[1] Ruffin defines the second best optimum in terms of a given consumption distortion. It will be shown, however, that the optimality of a zero effective tariff rate is not unambiguous in terms of a given *production* distortion and that, in fact, there exist two different second best optima in this context.

Following Ruffin, it will be assumed that the home country is small and produces two final goods, X_1 (exportables) and X_2 (importables). The production of import-competing goods requires the use of an imported material input, X_3, which is not produced domestically and which is used in fixed proportions. Production functions are homogeneous of the first degree and quasi-concave, and all competitive assumptions are assumed to hold. To avoid income distribution problems, we postulate a collective utility function $U=U(C_1, C_2)$, where C_i is the aggregate consumption of the ith good. With p denoting the relative price of the second commodity in the home country, X_1 being taken as the numeraire, changes in real income or welfare measured in terms of the numeraire are given by[2]

$$dy = \frac{dU}{U_1} = dC_1 + pdC_2 \tag{1}$$

Let p^* denote the fixed foreign price ratio of final commodities, p_3 and p_3^* the domestic and foreign relative price of the intermediate good, respectively, and k the fixed intermediate input coefficient in the import-competing sector. Balanced trade then implies

$$E_1 + p^*E_2 + p_3^*kX_2 = 0 \tag{2}$$

where $E_i=C_i-X_i$ is the domestic excess demand for the ith good. From (1) and (2), the change in welfare for small changes in p, p_3, and X_i that satisfy the budget constraint is readily shown to be given by

$$\begin{aligned} dy = [&dX_1 + (p - kp_3)dX_2] \\ &+ k(p_3 - p_3^*)dX_2 \\ &+ (p - p^*)dE_2 \end{aligned} \tag{3}$$

Since along the production possibility curve the marginal rate of transformation between the two final goods equals the ratio of value-added, the first bracketed term on the right-hand side of the above equation is zero. On the other hand, if t and t_3 denote, respectively, the ad valorem tariff rates on imports of the final and the intermediate goods, then $p=p^*(1+t)$ and $p_3=p_3^*(1+t_3)$. It follows that

$$dy = kt_3p_3^*dX_2 + tp^*dE_2 \tag{4}$$

Substituting for dE_2, we now obtain

$$dy = -p^*(t-\theta_3t_3)dX_2 + tp^*dC_2 \tag{5}$$

where $\theta_3=kp_3^*/p^*$ is the distributive share of the material input at free trade prices.

This last expression for the change in welfare may be further simplified by noting that the domestic output of the importable good, X_2, is a function of p and p_3, and that

* Department of political economy, University of Toronto. I am indebted to James R. Melvin, J. Clark Leith, and an anonymous referee for their valuable comments on the initial draft of this paper.

[1] Ruffin also showed that the second best optimum rate of effective protection is zero when the imported input is used in the export sector or in both sectors provided another unbiased measure of effective protection which he proposes is adopted. However, we shall not be concerned with this problem here.

[2] The reader will recognize the technique used by Ronald Jones.

$(\partial X_2/\partial p_3) = -k(\partial X_2/\partial p)$.[3] It can thus be seen that

$$(6) \quad dX_2 = \frac{\partial X_2}{dp} \cdot \frac{dp}{dt} dt + \frac{\partial X_2}{\partial p_3} \frac{dp_3}{dt_3} dt_3 = p^*(dt - \theta_3 dt_3) \frac{\partial X_2}{\partial p}$$

On the other hand, C_2 is a function of the commodity price ratio, p, and real income:

$$(7) \quad dC_2 = \frac{\partial C_2}{\partial p} \frac{dp}{dt} dt + \frac{\partial C_2}{\partial y} dy = p^* \frac{\partial C_2}{\partial p} dt + \frac{m_2}{p} dy$$

where m_2 is the marginal propensity to consume importables in the home country. Substituting for dX_2 and dC_2 in equation (5) yields equation (8).

$$(8) \quad dy = \frac{p^{*2}(1+t)}{(1+m_1 t)} \cdot \left\{ -(t - \theta_3 t_3) \frac{\partial X_2}{\partial p} (dt - \theta_3 dt_3) + t \frac{\partial C_2}{\partial p} dt \right\}$$

where

$$m_1 = 1 - m_2, \quad \frac{\partial X_2}{\partial p} > 0; \quad \text{and} \quad \frac{\partial C_2}{\partial p} < 0$$

Consider first the case analyzed by Ruffin in which a tariff on the final good is arbitrarily set ($t = \bar{t}$) and the problem is then to find the second best optimum tariff on the imported input, $t_{3,op}$. In this case, we can write

$$(9) \quad \left.\frac{\partial y}{\partial t_3}\right|_{t=\bar{t}} = \frac{p^{*2}(1+\bar{t})}{(1+m_1\bar{t})} \cdot \left\{ (\bar{t} - \theta_3 t_3) \frac{\partial X_2}{\partial p} \cdot \theta_3 \right\}$$

[3] As long as X_1 does not use the imported input, the output of X_2 is a monotonically increasing function of $(p - kp_3)$.

If t_3 is at its optimum value, $\partial y/\partial t_3 = 0$, so that

$$(10) \quad t_{3,op} = \bar{t}/\theta_3$$

Remembering that the Barber-Johnson-Balassa measure of effective protection is $ERP = (t - \theta_3 t_3)/(1 - \theta_3)$, it follows that for any given $\bar{t}$, the optimum effective rate of protection is

$$(11) \quad ERP_{op} = \frac{[\bar{t} - \theta_3(\bar{t}/\theta_3)]}{(1 - \theta_3)} = 0$$

which is the result derived by Ruffin.[4]

It must be noted that this result was derived under the assumption that the first best optimum solution (namely, free trade) is not attainable because of the presence of a nominal tariff on the final commodity. The question arises, however, as to whether the second best effective tariff rate is zero when the initial distortion is a nominal tariff on the imported intermediate good. Let $\bar{t}_3$ be any such arbitrarily chosen tariff; the effect of imposing a tariff on the final commodity is now given by

$$(12) \quad \left.\frac{\partial y}{\partial t}\right|_{t_3=\bar{t}_3} = \frac{p^{*2}(1+t)}{(1+m_1 t)} \cdot \left\{ -(t - \theta_3 \bar{t}_3) \frac{\partial X_2}{\partial p} + t \frac{\partial C_2}{\partial p} \right\}$$

It follows that the second best nominal tariff on the final good is

$$(13) \quad t_{op} = \frac{\theta_3(\partial X_2/\partial p)\bar{t}_3}{(\partial X_2/\partial p) - (\partial C_2/\partial p)} \gtrless 0 \quad \text{if } \bar{t}_3 \gtrless 0$$

and the second best effective rate of protection is

$$(14) \quad ERP'_{op} = \frac{\theta_3(\partial C_2/\partial p)\bar{t}_3}{(1-\theta_3)[(\partial X_2/\partial p) - (\partial C_2/\partial p)]} \lessgtr 0 \quad \text{if } \bar{t}_3 \gtrless 0$$

In other words, *given any arbitrary tariff (subsidy) on the material input, the second best effective tariff rate is negative (positive)*!

[4] See his equations (5) and (9), pp. 264–65.

This result stands in sharp contrast with Ruffin's and is indeed odd. However, the intuitive explanation of this asymmetry between the two cases is quite simple. The initial imposition of a tariff on imports of the final good results in a production loss as well as a consumption loss. When a tariff on the imported input is then levied, the effect of the latter is to reduce the production loss as the output bundle produced moves toward its free trade composition, while the original consumption loss remains unaffected. When the effective tariff rate falls to zero the production loss will have been totally eliminated and further increases in t_3 will again result in production losses.[5]

On the other hand, the initial imposition of a tariff on the intermediate good involves only a production loss. If a tariff on the final commodity is subsequently introduced, this production loss is reduced but a consumption loss is simultaneously created. Thus, it is no longer clear that the attainment of the second best optimum involves raising the tariff on the final good up to the point where the effective rate of protection is zero, i.e., up to the rate at which the production bundle is the same as under free trade. In fact, the optimum *ERP* will always be negative in this case[6]—expect in the event that the two final goods are consumed in fixed proportions ($\partial C_2/\partial p=0$) or if $\partial X_2/\partial p$ approaches infinity. Conversely, if the import-competing sector is being protected through a subsidy towards the use of the material input, the second best policy involves an output tariff which is not large enough to wipe out the protection given to that sector.

Thus, it turns out that we cannot in general conclude that the second best rate of effective protection is zero in the small country case. This will be true if the exogenous distortion takes the form of nominal protection of the import-competing industry, but not if the distortion is a tariff or subsidy on the imported input used by that industry. This, in turn, casts some doubt on Ruffin's definition of an *efficient* measure of effective protection: " . . . if when the effective tariff on the imported final good is zero and the tariff on the final good is non-zero, a second best optimum is achieved, then the effective tariff is said to be *efficient*" (p. 265). While Ruffin argues that the Barber-Johnson-Balassa measure is efficient, at least for the case where the commodity using the imported material input is itself an importable good, we have shown that this definition holds only if additional qualifications are made, namely, a specification of the nature of the initial distortion. Indeed, it appears that no measure of effective protection can be devised that will be unambiguously "efficient" in this sense and the concept of efficiency is thus of limited use for practical policy making.

[5] This brings out the point—not always explicitly recognized—that there are production losses from too much specialization as well as from too little.

[6] The reader may wonder whether it is not possible for the additional consumption losses from successive increases in t to be so small as to make it profitable to raise the latter to the point where the $ERP=0$. However, we must note that reductions in production losses are proportional to the effective rate of protection $(t-\theta_3\bar{t}_3)$ while increases in consumption losses are proportional to the nominal tariff rate t. Thus, for any non-extreme values of $\partial X_2/\partial p$ and $\partial C_2/\partial p$, it will always be true that for t sufficiently close to $\theta_3\bar{t}_3$, $-(\partial X_2/\partial p)\cdot(t-\theta_3\bar{t}_3)>-(\partial C_2/\partial p)t$.

REFERENCES

R. W. Jones, "International Capital Movements and the Theory of Tariffs and Trade," *Quart. J. Econ.*, Feb. 1967, *77*, 1–38.

R. J. Ruffin, "Tariffs, Intermediate Goods, and Domestic Protection," *Amer. Econ. Rev.*, June 1969, *59*, 261–69.

Industrial Prices, as Administered by Dr. Means

By GEORGE J. STIGLER AND JAMES K. KINDAHL*

The most fundamental proposition in the administered price literature is that prices in industrial markets, especially those which are oligopolistic in structure, are unresponsive to changes in general business conditions, and that this behavior is pervasive. Our study, *The Behavior of Industrial Prices*, is an examination of the evidence regarding this view.

Our central purpose was to ascertain whether "administered prices" which, as reported by the Bureau of Labor Statistics (*BLS*), were essentially unresponsive to declines of general business conditions, were in fact unresponsive: that is, our central task was to obtain transaction rather than quoted prices. To that end, we collected buyers' prices on commodities that met two conditions:

i. The commodities were those which did not undergo large quality changes over cycle-length periods of time (so the prices would be meaningful), and

ii. The prices were generally considered to be administered.

It follows that we were making a one-way test of the existence of administered prices: we deliberately omitted all industrial prices where the phenomenon of administration had not been urged with some plausibility. This concentration on administered prices had only three possible (but by no means certain) exceptions of which we were aware: bituminous coal, wood flooring, and plywood, where excellent price series became available.

The concentration of the sample in the area of administered prices means, of course, that we were not searching the universe of industrial prices to determine the extent of "administration" but rather that we were examining the behavior of so-called administered prices. Even if 100 percent of our sample showed a particular type of perverse price behavior, that would leave open the question of how the remainder (81.1 percent by weight) of nonagricultural and nonfood wholesale prices behaved.

Gardiner Means has reinterpreted the evidence on pricing behavior presented in our book and he is duly rewarded for his diligence by finding that "the administered-price thesis" is fully confirmed (1972). We shall restrict our comments to 1) the criteria of administered prices, and 2) several general questions of empirical procedure; these are the basis for Means' dissent from our conclusions.

I. Means' Concepts of Price Administration

Means has great claims to initiating the literature of administered prices (although F. C. Mills' claims are not negligible), but it is not his personal prerogative to designate which prices are administered in the eyes of the economics profession. The very success of Means' writings—and we reckon him among the most influential of economists in the history of this country—implies that they no longer have any special meanings he might wish to attach. For scientific purpose, it is the accepted view of administered prices that must prevail. Most economists specializing in industrial organization will accept our list of administered prices as generally valid (although incomplete).

Of course we concede Means' right to define administered prices for himself as he wishes, and we would be glad to segregate out a subset of prices and label them "Means-administered," but this proves to be virtually impossible—at least for us. It will be useful to comment upon his changing criteria of administration.

Initially his definition of administered prices was as follows:

> A market price is one which is made in a market as the result of the interaction of buyers and sellers. The prices of

* University of Chicago and University of Massachusetts, respectively.

> wheat and cotton are market prices. . . .
>
> An administered price is essentially different. It is a price which is set by administrative action and held constant for a period of time. We have an administered price when a company maintains a posted price at which it will make sales or simply has its own prices at which buyers may purchase or not as they wish. [1935, p. 1]

This distinction is palpably vague. Even on an organized exchange a buyer or seller may have a bid or ask price which persists for a long time. Conversely, what if one of numerous firms in an industry changes prices each month, but it is a different firm each month—the *BLS* data often display this amazing behavior—is the *market* price administered? Or if the firm makes different offers for forward contracts of different length—is its price constant over time?

In his actual practice during the 1930's, Means defined administration as infrequency of price change, and measured it simply by frequency of price change per price reporter. Thus, if three sellers of a product each made one price change in a three-year period, the price flexibility measure would be 1/35 whether the changes were in the same month or different months. The literature of economics did not faithfully follow Means in this respect; it introduced two different criteria which are absent from his early work:

i. Oligopoly or high concentration was identified with administration; Means expressly disavowed this connection.[1]

ii. Frequency of price change received less emphasis, partly as a result of the increased attention to concentration, in favor of amplitude of price change.

When Means applied himself to inflationary conditions in the 1950's, he modified his criteria in several respects:

i. He necessarily used the inappropriate *BLS* frequency of price change, whose behavior is strongly influenced by the number of price reporters.[2] Hence, measured frequency of price change became heavily dependent upon the number of price reporters.

ii. He modified the classification resting upon frequency of price change in a highly subjective fashion:

> The classification of price groups into those dominated by price administration, mixed, and those dominated by market prices is based upon general knowledge. It is consistent with the findings of the BLS on average frequency of price change except in the five cases [out of a total of 15] as is shown in the table below: . . . [1959, p. 4756].

Here Means was dealing with categories of *BLS* prices rather than individual prices, but the criteria should not be affected.

iii. A tacit acceptance of the congruence of his classification with that of competition and monopoly became apparent, although not wholly explicit. Thus, he refers to "classically competitive prices for food, leather, lumber, textiles, and farm products . . . " (1959, p. 4758, also p. 4775). In the same oral testimony, he says, "I know well that the rubber industry, the manufacturing of rubber, is fairly concentrated. Therefore we could appropriately classify even this exception [the rubber and rubber products industry] as black" [i.e., as an administered price industry; a chart presented had colored industries said to have administered prices in black] (1959, p. 4758).

In his most recent article, Means offers still another criterion of administered prices. He omits fifteen commodities as being market-dominated. Eight copper related prices, for example, are deleted because they are "fabricated from highly competitive raw materials" (1972, pp. 295–96, n.). This is an almost new criterion of nonadministration, wholly absent from his works in the 1930's,

[1] Means, (1935), p. 1: "Administered prices should not be confused with monopoly. The presence of administered prices does not indicate the presence of monopoly. . . . In many highly competitive industries, such as the automobile industry, prices are made administratively and held for fairly long periods of time."

[2] Earlier, he had access to individual price reporter data, and our failure to realize this led to the inappropriate criticism which was withdrawn in the *Journal of Economic Literature*, Sept., 1971, p. 852. The data of the 1953–56 study of the *BLS* are fully vulnerable to these criticisms, see Stigler and Kindahl, pp. 18–20.

but with partial anticipation in his testimony before the Kefauver Committee. Yet even in Means' own terms the omissions are odd: The *BLS* price of copper ingots shows in our 10-year period, subperiods of 33, 13, and thrice of 7 months of price constancy.[3] Aluminum ingot prices behave similarly, though the rigidity is less dramatic.

He also has two new interpretations of the implication of price administration:

1. He now believes that administered prices should be corrected for trend (1972, p. 299). This is a novel departure from previous practice—with important implications.

2. " . . . the administered-price thesis is not concerned with regular seasonal price jumps" (1972, p. 301). This again is a wholly new interpretation.

Means' theory has indeed become difficult to refute or confirm. In the absence of a well-defined set of criteria, and rigorously derived implications, no test can be convincing. Means' criteria have changed—or, perhaps, evolved—over time. The implications have become so broad as to be almost uselessly vague: "Basically, the administered-price thesis holds that a large body of industrial prices do not behave in the fashion that classical theory would lead one to expect." His empirical work using our data embodies this statement without further qualification. We think that virtually all of our price series are appropriate to what economists in general mean by administered prices, and that our interpretation of its implications conform to an important segment of thought on the question. We think that Means' extreme flexibility of concept of price administration does his work a disservice.

A final word on Means' "theory." He has never offered an acceptable theoretical explanation for administered prices (=infrequently changed price quotations) as either a phenomenon of individual business or as a market phenomenon. We provide an orthodox explanation of long-term contracts by individual businesses in our book, and find no clash between traditional theory—once transaction costs are introduced—and observed behavior. Means insists upon limiting "classical" price theory to the large-number auction of a perishable good, and much of his frustration with the theory comes from this narrow vision.[4]

II. The Interpretation of Evidence

Means' paper comes to the conclusion, startling at least to us, that our new price data really indicate the opposite of what we assert: the new data, indeed, seem to him to provide overwhelming support for his administered-price thesis. Thus, his Table 1 retabulates our data to suggest that well over half our price series conform to his "truncated administered-price hypothesis." How can our statistics be truly so obliging to Means' purposes? There are three reasons.

First, he excludes nearly one-fourth of our price series as "raw material dominated" (a concept left undefined). We have already argued that this exclusion is mistaken.

Second, he redates the turning dates of cycles. We adopted the reference cycle dates of the National Bureau of Economic Research (*NBER*)—a chronology which has been widely accepted by the economics profession. For reasons best known to himself, Means constructs new cycle dating by using only his interpretation of the Index of Industrial Production; this gives the same results as the *NBER* dating for the first contraction (July 1957–April 1958) and for the beginning of the first expansion (April 1958); he dates the end of the expansion at June 1959, eleven months short of the May 1960 peak of the *NBER* series. Means dates the second contraction as beginning in January 1960. Thus, he leaves the last six months of 1959 in the limbo of neither expansion nor contraction—a procedure totally foreign to the accepted methods of defining cycles. Finally, he finds the expansion ended in March 1962. (The *NBER* reference cycle expansion does not begin until January 1961).

[3] The Houthakker report on copper prices states that with the pegged domestic price producers "have traditionally attempted to mitigate fluctuations in copper prices by changing prices only at discrete intervals. . . ." (Council of Economic Advisers, p. 9).

[4] Means asserts a clash because he has a rigid, elementary competitive price model (1972) p. 302.

The expansion beginning in 1961 (as seen by the *NBER* reference cycle) is the longest in the history of the United States. We noted that taking the entire period of 1961–66 as an expansion would give so much weight to trend factors as to be useless as a cyclical measure. Somehow, we had to choose part of that expansion for analysis. Means wishes to choose the first 14 months; we chose the last 24. Our grounds were that in the 1961–66 period, a number of indicators (*GNP*, constant dollar *GNP*, industrial production, expenditures on new plant and equipment) showed higher annual percentage increases in the last two years of the period than in earlier years of the expansion. Moreover, inflation began in these latter two years, and we wished to examine both the accuracy of the Wholesale Price Index (*WPI*), and the behavior of "administered" prices, during this period of rising prices.

Third, Means enlarges the predictions for price behavior of the administered-price hypothesis to include *any* departure from strong classical predictions, and his tables are so labelled. Let us rewrite Means' Table 1 with neutral headings breaking the heading "Conforming to the Truncated Administered-Price Thesis" into two parts.

TABLE 1—MEANS' TABLE 1 REARRANGED AND DISAGGREGATED

	1st Contraction July 57–Apr. 58	1st Recovery Apr. 58–June 59	2d Contraction Jan. 60–Jan. 61	2d Recovery Jan. 61–Mar. 62
Price Increased	15	17	12	5
No Change In Price	19	10	14	16
Price Decreased	16	23	24	29
	50	50	50	50

Consider the two recoveries: the decreases of prices in the recoveries account for *more than half* of the 100 observations on Means-defined administered prices in recoveries. These are classified in his tables as "conforming to the truncated administered-price hypothesis." Shall we call this "administered-price deflation"? The extraordinary intellectual imperialism of this label should be clear.[5] Firms using their market power to force prices down during an expansion are not displaying "administered price behavior" to the profession. We are hardly ready to concede this price behavior to be telling evidence in favor of the view that prices of such goods do not fall in contractions and rise only tardily in expansions.

In the two brief contractions allowed by Means, one shows about a third of the prices declining, the other nearly one-half. Considering the shortness of the period (eight and twelve months, respectively) compared with the typical length of contracts used by our buyers (twelve months) we do not find this convincing evidence of price administration—especially since the fifty commodities surviving in the table are all certified to have administered prices.

We shall not subject the reader to our more detailed differences from Means, such as the overrepresentation of steel products in our sample (which seems evident on the basis of *WPI* weights), or their atypical behavior (which seems evident from the charts in our book), or our use of weighted averages in summarizing experience of two cycles. The *facts* on price behavior are at the reader's disposal and he can form his own judgment.

III. Conclusion

The most fundamental proposition in the administered price literature, to repeat, is that industrial markets, especially those which are oligopolistic in structure, are unresponsive in their pricing to changes in general business conditions. Our work seriously undermines this proposition.

[5] Means goes still further: he finds some commodities with a downward trend in price, whose rates of fall in contractions were less than in expansions. By using deviations from trend (shown as "data corrected for trend"), rather than raw data, he converts them to commodities whose prices "rise" (relatively) in recessions, fall in contractions—and then claims these as behaving as administered-price theory predicts! (see Means (1972) p. 299).

The standard theory of pricing shared by virtually all modern economists has serious limitations and unresolved problems. In our study, we lamented our tardy recognition of the fascinating problem of determining the nature of the contracts under which most industrial goods are sold. We analyzed prices in specific cycles, to get wholly obscure results. Cleverer men will find difficulties with the received theory that never occurred to us. But we do not believe that these known and unknown difficulties with received theory are favorable to Means. His theory, if it be a useful one, cannot be, as at times he believes:[6] prices behave differently than most economists believe—that theory has, no doubt, ample content but no direction. He once observed, no doubt whimsically, that "... if one introduces a single administered price into the [Walrasian] system, the equations cannot be solved" (1962, p. 28). The economy is not all that fragile; a single administered price—let it behave as insanely as the Devil can devise—will have a negligible influence. The thesis that modern economics has received from Means, whatever he intended, is that perverse or unresponsive price behavior is widespread. Our study, without benefit even of data from a large or protracted recession, contradicts that thesis.

[6] An amazing example of this Phoenix property of Means' theory is the discussion of "specific cycles"—large changes in output of individual industries. He views as strong confirmation of his thesis the fact that we found very few price increases in these large expansions (there were few sufficiently large contractions in our period). Yet two pages later he asserts that according to his thesis, prices should *rise* in periods of sustained expansion of output! "... A finding of a predominent tendency for prices to rise in this period could not possibly support either thesis as against the other" (1972, p. 299).

REFERENCES

G. C. Means, *Industrial Prices and Their Relative Inflexibility*, U.S. Senate Document 13, 74th Congress, 1st Session, Washington 1935.

———, *Hearings on Administered Prices, Part 9*, Senate Subcommittee on Antitrust and Monopoly, Washington 1959, pp. 4745–60.

———, *The Corporate Revolution in America*, New York 1962.

———, "The Administered-Price Thesis Reconfirmed," *Amer. Econ. Rev.*, June 1972, *61*, 292–306.

G. J. Stigler and J. K. Kindahl, *The Behavior of Industrial Prices*, New York 1970.

Council of Economic Advisers, *Report of the Subcommittee on Copper*, Washington, May 18, 1970.

Environmental Externalities and the Arrow-Lind Public Investment Theorem

By ANTHONY C. FISHER*

In a recent article concerning the evaluation of public investments Kenneth Arrow and Robert Lind have shown that, as the net returns to an investment of given size are shared by increasingly many individuals, the risk premium for the representative individual vanishes and, more importantly and perhaps surprisingly, the aggregate of these premiums for all individuals also approaches zero.

The purpose of this note is to consider whether, and how, this result and the implied policy of risk-neutrality in the evaluation of public investments might have to be modified where these investments produce uncertain environmental side effects. In particular, suppose that in addition to the direct costs of undertaking a project, already netted out of the returns, there are other, external costs in the form of environmental damages, borne perhaps by a different set of individuals than those who benefit from the project. In what follows, conditions under which this would affect the Arrow-Lind results are defined, and the results appropriately modified. The note is concluded with a discussion of the empirical relevance of key assumptions in the analysis, and some implications for policy.

I

Following Arrow and Lind, let the welfare function for each of n identical individuals be written

$$(1) \qquad W(s) = E[U(A + s\overline{B} + sx)]$$

where welfare W is the expected utility $E[U]$ of (uncertain) income A independent from the particular public investment under consideration, plus the share s of (net) income B from the investment, made up of the expected returns $\overline{B}$ plus a random component x with mean zero. If all share equally in the net returns, then $s=1/n$. Assuming the representative individual is risk averse, there exists a number $k(n)>0$ such that

$$(2) \quad E\left[U\left(A + \frac{\overline{B}}{n} + \frac{x}{n}\right)\right] = E\left[U\left(A + \frac{\overline{B}}{n} - k(n)\right)\right]$$

i.e., the individual is indifferent between accepting the risk x/n and paying the amount $k(n)$.

Arrow and Lind note first that $\lim_{n\to\infty} k(n) = 0$, and then prove that $\lim_{n\to\infty} n\, k(n) = 0$, which is not intuitively obvious. The meaning of this result is that if n is large enough, then only the expected value of net benefits should be taken into account in evaluating the investment—even though each of the affected individuals is assumed to be averse to risk.

Note that costs and returns are finite, and sum algebraically to B. Any part of B that goes to one individual does not go to another, so that as B is (equally) divided among increasingly many individuals, the size of the share going to each decreases. But suppose, as has been suggested, that in addition to the costs already included in B, the investment generates environmental damages, for example air pollution, which can be considered a "public bad" in the sense that what is received by one economic unit does not appreciably reduce what is received by others. To put the matter a bit differently, let $C(m)/m$ be the environmental damages borne by each of m identically affected individuals. Then $dC(m)/dm>0$, i.e., total damages $C(m)$ are increasing with m, per-

* Resources for the Future, Inc. I have benefitted from conversations with Kenneth Arrow, John Brown, and Harl Ryder, and from comments on an earlier draft by Barry Chiswick, Donald Hester, John Krutilla, Howard Kunreuther and Kerry Smith.

haps sufficiently rapidly that $C(m)/m$ remains constant, or even increases, with increase in m.[1] In this not untypical situation, clearly, the stake in the outcome of each of the affected individuals, and therefore also the cost of risk to each, does not decrease with increases in their number—or, of course, in the number of those who benefit.

The implications of this for the cost of risk in the aggregate, and in turn for the social evaluation of the investment, are spelled out in the following exercise within the Arrow-Lind framework.

Suppose, then, there are n individuals who share equally in the net benefit B from an investment, and m (different) individuals who bear equally the environmental damages. An efficient policy is one that maximizes

$$(3) \qquad nE\left[U\left(A + \frac{B}{n} - \frac{\pi}{n}\right)\right]$$

subject to

$$(4) \quad mE\left[U\left(A - \frac{C(m)}{m} + \frac{\pi}{m}\right)\right] \geq mE[U(A)]$$

where π is the total compensation that could be transferred from the n gainers to the m losers and all other symbols are as defined earlier.

Dividing $C(m)$ into two parts, $\overline{C(m)}$ expected costs, and $Z(m)$, a random part with mean $\overline{Z(m)}=0$, (4) can be rewritten as

$$(5) \quad E\left[U\left(A - \frac{\overline{C(m)}}{m} - \frac{Z(m)}{m} + \frac{\pi}{m}\right)\right] \geq E[U(A)]$$

Assuming the representative individual who suffers the damages is risk averse, there is some positive amount he would be willing to pay in order to remove the risk represented by $Z(m)/m$. At the margin,

$$(6) \quad E\left[U\left(A - \frac{\overline{C(m)}}{m} - \frac{Z(m)}{m} + \frac{\pi}{m}\right)\right] = E\left[U\left(A - \frac{\overline{C(m)}}{m} - l(m) + \frac{\pi}{m}\right)\right]$$

where $l(m)$ is the positive amount which would leave him indifferent between paying and accepting the risk. Combining (5) and (6), we have

$$(7) \quad E\left[U\left(A - \frac{\overline{C(m)}}{m} - l(m) + \frac{\pi}{m}\right)\right] \geq E[U(A)]$$

which implies

$$(8) \qquad \pi \geq \overline{C(m)} + ml(m)$$

Since, however, (3) is to be maximized, π must be as small as possible, so that

$$(9) \qquad \pi = \overline{C(m)} + ml(m)$$

Alternatively, B should be undertaken (assuming no constraint on capital) if

$$(10) \qquad \overline{B} - \overline{C(m)} - ml(m) > 0$$

In the absence of individual risk aversion, the cost of risk-bearing $l(m)$ vanishes, leaving the Arrow-Lind result that a project should be undertaken if the expected value of benefits net of all costs is positive. Where environmental losses are collectively borne, and the risk to those who bear them remains significant, then, this result must be adjusted. To the cost of the project must be added not only the expected value of these losses but also the amount that would be required to compensate those who bear the risk. Although information on these magnitudes would ordinarily be difficult to come by, it is at least clear that some adjustment for risk ought to be made.

The result expressed in equation (10) can be generalized rather simply to treat all individuals symmetrically. Proceeding as before, write

[1] As a matter of fact, the analysis that follows requires only that $\lim_{m\to\infty} C(m)/m \neq 0$. One way in which this condition could fail to hold, i.e., in which $C(m)/m$ could approach zero, would be if a variety of other adverse effects of a very large increase in m, not related to the investment externality, were to render the separate effect of the investment negligible.

$$(11) \quad E[U_i(A_i + \overline{B}_i + x_i)] = E[U_i(A_i + \overline{B}_i - k_i)]$$

where U_i is the utility of the ith individual, from preinvestment income A_i, expected investment benefits $\overline{B}_i$ *net* of environmental costs, and a random investment return x_i. Equation (11) says that there is some positive amount k_i the risk-averse individual would pay that would leave him indifferent between paying and accepting the risk represented by x_i. Let T_i be the amount, positive or negative, that could be transferred from each individual and leave him as well off after the project is undertaken as before. In symbols,

$$(12) \quad E[U_i(A_i + \overline{B}_i - k_i - T_i)] = E[U_i(A_i)]$$

Then

$$(13) \qquad T_i = \overline{B}_i - k_i$$

and the project should be undertaken if

$$(14) \qquad \sum_i T_i = \sum_i (\overline{B}_i - k_i) > 0$$

II

One important assumption underlying these conclusions is of course that the environmental damages associated with the investment represent a nonnegligible fraction of the real income of the affected individuals. Were this not the case, i.e., were they in fact negligible, clearly it would matter little that they are not reduced to individuals as they are spread. This suggests that the appropriateness of the adjustment to net benefits for aversion to the risk of environmental damage may be an empirical question. In at least some cases, as where significantly increased mortality rates, or reduced residential property values have been attributed to air pollution, the environmental effects of an investment project may well be large enough to make the adjustment appropriate.[2] In particular, effects that are cumulative, or irreversible, may fall in this category.[3]

Another important assumption is that the risk is not (costlessly) transferable from the affected individuals to the larger community. If private markets for transferring the risk exist, then the Arrow-Lind results follow, as indeed they demonstrate in the first part of their paper—though not with reference to "collective" risk of the sort considered above. The second part, on which this discussion has been based, is designed explicitly to demonstrate the validity of their results precisely when these markets do not exist. The analysis of the preceding section might then be interpreted as an argument for government provision of insurance, where it is not privately available, against the risk of environmental damage.

The difficulty with this approach is that where markets for transferring risk do not exist, presumably this is because the moral hazard is too great, or the cost of making the complicated contracts that would be required is too high, as Arrow and Lind also note. The case for government provision of insurance against the risk of environmental damage then depends in part on the ability of government to reduce the moral hazard, or the transaction costs. One advantage government has, of course, is that it can in effect enroll the entire community in an insurance program, thereby maximizing risk spreading, or what is the same thing, minimizing the cost of risk bearing, as in the Arrow-Lind model. Thus suppose all costs (as well as all benefits) associated with a particular project, including any environmental damages, were made internal to the project in the sense that returns net of all costs were shared equally. In the framework of the analysis of the preceding section, the m affected individuals would not alone bear the cost and the risk, which would be spread, with the benefits, to all members of the community.

A policy to internalize all of the costs of a public project, including in particular the cost of environmental risk, is certainly appealing, and is consistent with economists'

[2] For empirical studies of the health and property losses attributable to air pollution, see Lester B. Lave and Eugene P. Seskin, and Ronald Ridker.

[3] For a discussion of the special implications of irreversible changes for the problem of environmental preservation, see Kenneth Arrow and Fisher.

recommendations for the control of privately generated external costs. The difficulty here again is that an attempt to transfer risk in this fashion is likely to be accompanied by substantial transaction costs, broadly conceived. Suppose, for example, we are considering whether to build a jetport near enough to a residential neighborhood that some people are likely to be affected by noise and air pollution. In order to make these uncertain environmental costs internal to the project in the sense suggested above (and as they have generally not been), costs of education, persuasion, or even coercion might be incurred. This is because some other people (those not in the affected area) presumably benefit from the project so long as the environmental costs are *not* internalized.

Or suppose the question is whether to construct a dam for purposes of power generation, where this construction necessarily results in the flooding of a canyon, obscuring the walls and converting a free-flowing river into a pool. Suppose, further, that river and canyon, if left undeveloped, are capable of providing valuable recreational and scientific services (though perhaps not to the same individuals who would benefit from the dam). It may be difficult, i.e., costly, even for the government, to identify all those who lose as a result of the development, and to measure all of the losses, as would be required under any scheme that seeks to insure equal shares in the gains from the project net of all losses. What these examples suggest is the possibility that the cost of transferring the risk of some environmental damages from private individuals to the larger community may exceed the cost to these individuals of bearing the risk, where the damages would be hard to identify and measure, or are unequally distributed.

As a final point, note that the suggested transfer will be effective in reducing the cost of risk only if m is small relative to n, assuming again that the environmental costs to each of the m affected individuals are not reduced with increases in m.

What are the implications for policy of the several points raised in the note? First, it would be desirable to internalize, in such a fashion that they are shared more or less equally, the environmental costs of a public project. This is so not only for the traditional reason for internalizing all costs of a private project, namely to correct the distortion in allocation that would result from a failure to internalize, but also for another reason: to eliminate the deadweight loss that would result from privately borne risk of environmental damage. Second, the attempt to transfer this risk to the larger community may itself carry costs; and third, where these (transaction) costs are large enough that transfer is not profitable, the net benefit from a public project should be adjusted to take account of the attitude of the affected individuals toward the risk of environmental damage.

REFERENCES

K. J. Arrow and A. C. Fisher, "Environmental Preservation Uncertainty, and Irreversibility," *Quart. J. Econ.*, forthcoming.

K. J. Arrow and R. C. Lind, "Uncertainty and the Evaluation of Public Investment Decisions," *Amer. Econ. Rev.*, June 1970, *60*, 364–78.

L. B. Lave and E. P. Seskin, "Air Pollution and Human Health," *Science*, Aug. 21, 1970, *169*, 723–33.

R. Ridker, *Economic Costs of Air Pollution*, New York 1967.

A Note on Proportionally Distributed Quotas

By Rachel McCulloch and Harry G. Johnson*

Traditional theory has usually condemned trade restriction as an irrational or misguided interference with efficient resource allocation. The two notable exceptions have been the "optimum" tariff and the infant industry argument for protection. More recently, however, trade theorists have begun to investigate the possibility that tariffs are used to achieve certain "noneconomic" objectives, that is, ones reflecting considerations not captured by individual private consumption.[1] Their work has evaluated the relative efficiency of tariffs, production subsidies, and consumption taxes as instruments for implementing such objectives as a given increase in domestic production or restraint of imports below the free trade level.[2] This note seeks to extend the analysis by comparing a tariff (or equivalent quota) with a commonly used protective arrangement which we term a "proportionally distributed quota," as alternative means of pursuing various objectives of trade restriction.

We do not intend to de-emphasize the cost, even when minimized, of achieving certain objectives through protection. Rather, to the extent that these objectives are the end result of a process of political bargaining, they may be difficult to alter merely through demonstration of their welfare cost to the economy as a whole. Of course, calculation and promulgation of this cost makes it harder for special interest groups to obtain economic privileges at the expense of the majority. However, we should not overlook the possibility that certain noneconomic objectives are truly public goods which provide sufficient collective utility to the majority to justify their cost in terms of foregone consumption. In either case, it is desirable to provide the least-cost method of achieving the objective. For purposes of this paper, noneconomic can best be understood as "outside the model."

It is well known that any protective goal can be achieved through some combination of a consumption tax and a production subsidy. However, tariffs and quotas are employed much more frequently than direct subsidies—probably for reasons of political feasibility. In this paper we use a familiar model to show that a proportionally distributed quota implies a smaller tax and larger subsidy than a tariff or conventional quota, and thus provides a more economically efficient—but equally practical—means of achieving some objectives.

As long as markets remain competitive, import quotas distributed in most of the usual ways[3] are analytically equivalent to tariffs, since they will acquire a premium value in the market reflecting the domestic scarcity of the importable good.[4] The only difference is in the distribution of the proceeds. In fact, the import licenses can be auctioned to the public, in which case (assuming again that markets remain competitive and in particular that the supply of licenses is not monopolized) the distribution of the proceeds is also equivalent, since the auction should yield the same revenue as would be produced by a tariff. We designate this type of restriction below as a tariff/quota.

In contrast, the proportionally distributed quota is a system in which import rights are assigned to producers or consumers *in proportion to* their individual production or consumption of the restricted good. We show below that *which* group receives the licenses is irrelevant to the resulting equilibrium, for a given objective of protection (again, markets remaining competitive). This is analogous to the fact that the economic incidence

* University of Chicago.

[1] See, for example, Harry G. Johnson (1960).

[2] For instance, see Jagdish Bhagwati and T. N. Srinivasan.

[3] That is, *other than* in proportion to consumption or production.

[4] Bhagwati gives an extended discussion of equivalence under alternative market conditions.

of an excise tax or subsidy does not depend upon its legal incidence. Examples of proportionally distributed quotas are the *U.S.* oil import program, in which quotas are (or were) assigned to refiners in proportion to their refinery throughput,[5] and the system of content protection, by which import duties on parts or finished products or both are waived if a certain percentage of sales is accounted for by domestic value-added.[6]

For simplicity, we assume competitive markets and abstract from the possibility of a quota-induced change in market structure. We also exclude optimum tariff considerations by assuming a perfectly elastic foreign supply curve for the protected good, and ignore the problems raised by divergence between private and social costs or benefits. These complications will be briefly discussed later. We consider four possible noneconomic objectives of trade restriction: (a) to achieve a given ratio (below the free trade level) of imports to domestic production (or, equivalently, consumption); (b) to reduce imports; (c) to increase domestic production; and (d) to reduce domestic consumption. The first of these objectives has increasingly come to be regarded by protectionists as a "fair" compromise between the desire to protect domestic industry from foreign competition and the claims of foreign producers to access to domestic markets (or, possibly, fears on the part of exporters that trading partners may impose retaliatory restrictions). Most protective policies can be expressed in terms of one of these four alternatives.

In the accompanying figures, *DD* is the domestic demand curve for the protected good, *SS* the domestic supply curve, and P_f the world market price of imports. Figure 1 illustrates the standard cost of protection argument,[7] where a tariff or quota restricts imports to a level *FG*. We may define consumption cost to be *GHJ*, production cost to be *FKL*, and observe that tariff revenue (or proceeds from auction of import licenses) is *FGJK*. According to this analysis, the importing country is now worse off than under free trade. However, if we assume that the government behaves rationally, some further policy objective must underlie the restriction of trade, and this should be taken into account in evaluating the result. In terms of the four policy objectives suggested above, the protection imposed is consistent with: a) a desired ratio $k=XV/OX$ of imports to domestic production; b) a reduction of imports to *FG;* c) an increase of domestic production to *OX;* or d) a reduction of consumption to *OV*. We will compare the cost of achieving each of these objectives through a tariff/quota with that incurred when a proportionally distributed quota is used. The difference results from the fact that the proportionally distributed quota uses the value of the import licenses to influence directly the allocation of resources within the protected industry. In some instances, this reduces the degree of import restriction (and hence the welfare cost) required to achieve the desired result.

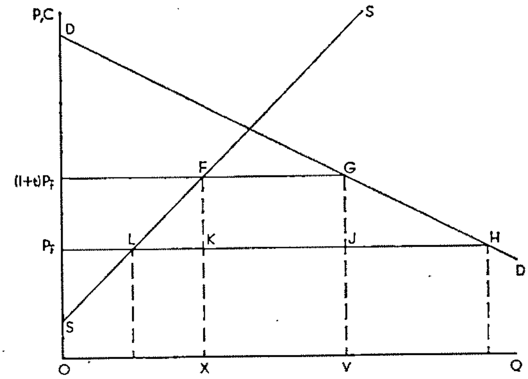


FIGURE 1. THE COST OF PROTECTION

Suppose that the policy objective is (a), a given ratio of imports to domestic production. Assume now that import licenses are distributed to each producer in proportion k to his individual output. This will cause domestic output (and hence imports) and consumption to increase above their levels under the tariff/quota used to achieve the same objective. To see why this occurs, con-

[5] For a detailed description of the mechanics of the program, see the Cabinet Task Force on Oil Import Control.

[6] Analyses of content protection are provided by Johnson (1963) and Bernard Munk.

[7] See, for instance, Charles P. Kindleberger, pp. 105–06.

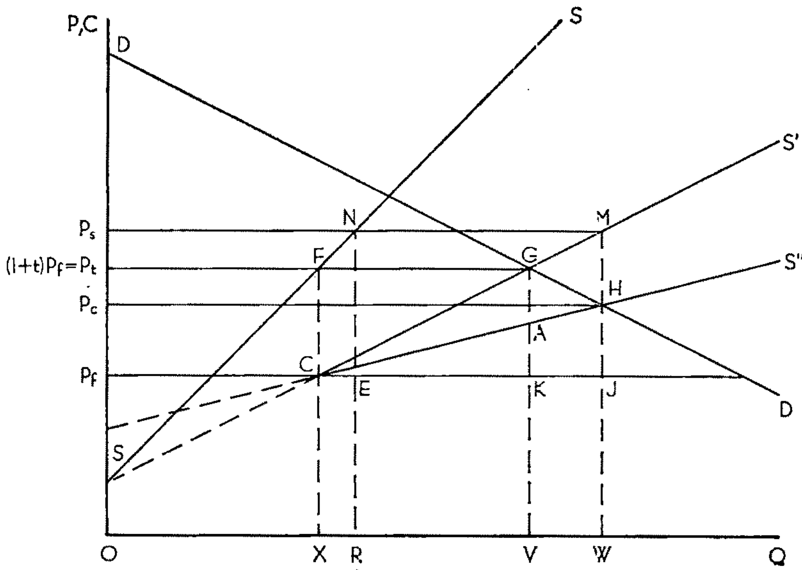


Figure 2. A Fixed Ratio of Imports to Domestic Production or Consumption

sider the position of producers at the tariff/quota equilibrium point, G, reproduced in Figure 2. The value of each license is $P_t - P_f$, the difference between the cost of importing an additional unit and the price which consumers are willing to pay for it. Receipts from an additional unit produced will be the market price P_t plus the value of the fractional import license thus earned, a total of $P_t + k(P_t - P_f)$. This exceeds the domestic producers' supply price, P_t. Therefore, production (and consumption) will be expanded until, at the new equilibrium point, H, domestic supply price P_s exceeds demand price P_c by exactly the value of the fractional import license earned by an additional unit of domestic output:

$$P_s - P_c = k(P_c - P_f)$$

Since consumption is equal to domestic production plus imports, the induced change in production is related to the increase in consumption by

$$(1 + k)(P_s - P_t)\frac{\Delta Q_s}{\Delta P_s} = (P_c - P_t)\frac{\Delta C}{\Delta P_c}$$

Use of a proportionally distributed quota results in a new relationship linking P_c, P_s, and P_f, which may be considered the *net* marginal cost or supply price of the producer-importer. An additional unit of domestic production, along with exercise of the import privilege thus earned, increases total cost by $P_s + kP_f$ and total supply by $1+k$ units. Hence, the net marginal cost of the protected good when supplied under the constraint of a fixed proportion of imports to domestic production is

$$\frac{P_s + kP_f}{1 + k} = \frac{1}{1 + k}P_s + \frac{k}{1 + k}P_f$$

This net supply relation is shown as the schedule S'' in Figure 2. The S'' schedule begins at point C of Figure 2, since for domestic prices less than P_f there would be no imports. As a consequence of competitive profit maximization, the net supply price must equal the market demand price, so that

$$P_c = \frac{P_s + kP_f}{1 + k}$$

i.e., the new equilibrium, H, is at the intersection of the S'' schedule with the domestic demand curve.

The S'' schedule (for domestic prices

above P_f) is the weighted average of foreign and domestic supply cost, the weights being the proportion of each in total consumption. To disaggregate this average, the S' schedule can be used. For every domestic supply price above P_f, the S' schedule shows $(1+k)$ times domestic output. For any level of total supply (consumption), the S'' schedule gives the net social cost while the S' schedule provides the domestic supply price of the required domestic output component of that total.

A proportionally distributed quota results in total consumption of OW of which OR is domestically produced at a supply price of P_s, while RW is imported at the world price, P_f. Total value of the import licenses, $(P_c - P_f)RW$, is equal to $(P_s - P_c)OR$, the excess of total cost of domestic output over its market price.

A tariff/quota is equivalent to a subsidy to production plus an equal tax on consumption. The proportionally distributed quota provides a different "mix" of the tax and subsidy elements. From Figure 2 we can see that a proportionally distributed quota results in an *additional* production subsidy of $P_s - P_t$ over that implied by the tariff/quota, plus a countervailing consumption subsidy of $P_t - P_c$, which partially cancels the consumption-taxing effect of the tariff/quota. The fact that a proportionally distributed quota implies a larger production subsidy and a smaller consumption tax than a comparable tariff/quota indicates that the former will be preferred whenever the consumption-tax aspect is an undesired side effect of achieving the policy objective.

From this, it can be seen that a proportionally distributed quota maximizes welfare subject to the constraint that imports are a fixed proportion of domestic production. The marginal condition for welfare maximization requires that price equal marginal cost. As previously indicated, the effective marginal social cost of the product when supplied under the given constraint is $(P_s + kP_f)/(1+k)$. By using the proportionally distributed quota, we equate this constrained marginal cost to the demand price of consumers. In Figure 2, the welfare gain which results from substituting this method for a tariff/quota is measured by $GHWV - FNRX - (KJWV - CERX) = GHA$. The area $GHWV$ is the gain from increased consumption of VW. The cost is $FNRX$ of additional domestic production and $KJWV - CERX = k(XR)P_f$ of additional imports. Net cost of additional consumption is given by the S'' schedule, so that the cost is also measured by $AHWV$. Thus, the welfare gain is $GHWV - AHWV = GHA$.

We stated earlier that the resulting equilibrium does not depend on whether consumers or producers receive the import licenses. This can be seen easily in terms of the net supply relation derived above. To maintain the same proportion of imports to domestic production, imports must be a fraction $k' = k/(1+k)$ of total consumption. Now consider the effective supply curve facing the consumer. For each unit he consumes, he obtains the right to purchase a fraction k' at the world price P_f. The remaining fraction $1-k'$ must be purchased domestically at the supply price P_s. Thus, the average price per unit is

$$k'P_f + (1 - k')P_s = \frac{k}{1+k}P_f + \frac{1}{1+k}P_s$$

Hence the net supply schedule facing the consumer is exactly that which is obtained when the licenses are distributed to producers. This symmetry does not depend on demand or supply elasticities.[8]

Policy objective (b), a given *level* FG of imports, is analyzed in Figure 3. The value of k is chosen so as to produce the desired equilibrium import level, FG $(=LR)$.[9] As discussed above, the value of the import licenses given to producers in proportion to their individual output results in increased domestic output and consumption. At the proportionally distributed quota equilibrium, R, the excess of total cost of domestic

[8] Although the two equilibria are alike in every *real* way, the domestic price P_c is not the same. Thus, the *nominal* value of the import licenses is also different.

[9] The appropriate value of k in this case is *less than* XV/OX, the proportion of imports to domestic production under a tariff/quota regime.

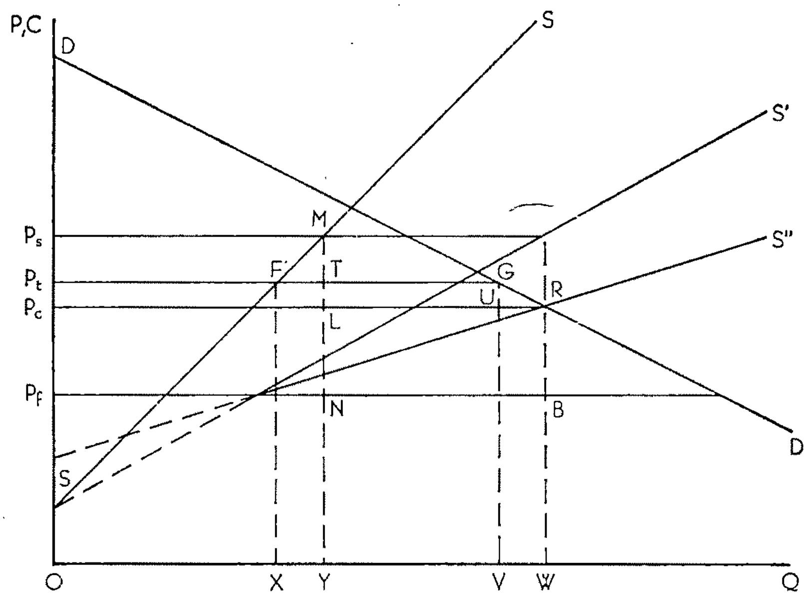


FIGURE 3. A FIXED LEVEL OF IMPORTS

production over total receipts from sale of that output (P_sP_cLM) is equal to the total value of the import licenses distributed ($LNBR$). But in contrast to case (a), in which imports expanded in proportion to domestic production, the increased consumption comes entirely from additional domestic production. This is because imports are fixed. Since consumers are willing to pay less (P_c) for an additional unit of output than it costs to provide (P_s) under the constraint of fixed total imports, the proportionally distributed quota is welfare-inferior to an ordinary tariff/quota in achieving this objective. The result follows because the objective is itself the limitation of trade. A tariff/quota is the optimal instrument in this case, since it equalizes the marginal cost of increased domestic production and reduced domestic consumption. The additional welfare cost incurred by use of a proportionally distributed quota is measured by $FMYX - GRWV = FMT + GRU$.

Policy objective (c), a fixed level of domestic production OX, is well suited to the use of a proportionally distributed quota; in contrast to the previous case, expanded domestic production is the stated objective of the program, rather than an undesired side effect of achieving a different end. The value of the import licenses is thus used to give a direct subsidy to the desired activity. There can be no efficiency loss due to higher domestic production with a proportionally distributed quota as contrasted with a tariff/quota, since production must be the same under either alternative. The increase in domestic consumption which results from the use of a proportionally distributed quota comes entirely from lower-cost foreign sources.

The value of k is chosen so as to result in domestic production of OX. As shown in Figure 4, the use of a proportionally distributed quota to achieve the desired production level results in a new equilibrium at N, with a consumer price of P_c and a net welfare gain of $GNMR$ over the use of a tariff/quota. Again, the excess P_sP_cHF of total domestic costs over sales is exactly equal to $HLMN$, the total value of the import licenses distributed. Changing from a tariff/quota to a proportionally distributed quota moves the economy toward the optimum, which would be achieved through the use of a direct production subsidy of $P_s - P_f$ per unit, with free trade.

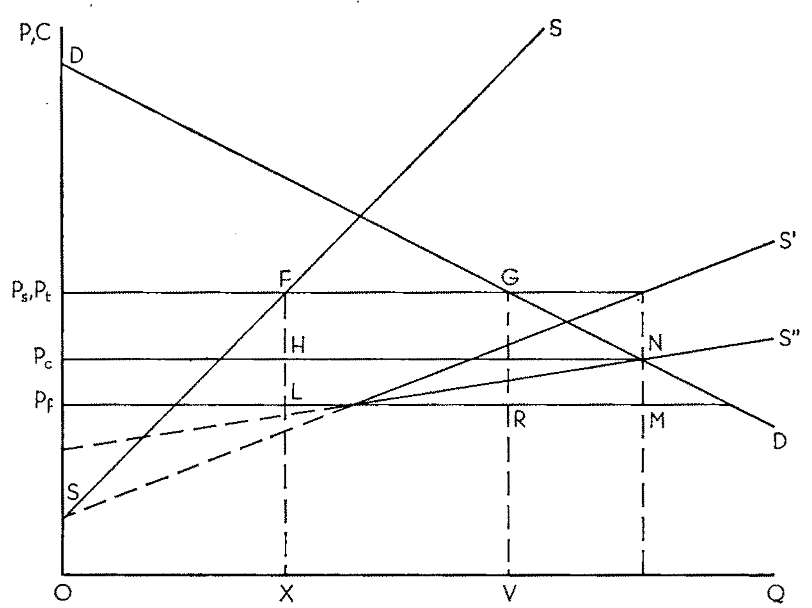


FIGURE 4. A FIXED LEVEL OF DOMESTIC PRODUCTION

For policy objective (d), restricting consumption to OV, the best instrument is a consumption tax. By analogy with the previous case, it is clear that the policy can be carried out more efficiently by a tariff/quota than by a proportionally distributed quota, since the former provides a higher consumption tax for a given level of production subsidy. The welfare loss from using a tariff instead of a consumption tax results from the implied subsidy to production, in this case an undesired subsidy. The additional subsidy produced by a proportionally distributed quota increases this welfare loss.

The analysis so far has assumed that the market supply and demand relationships reflect true social costs and benefits. Where a domestic distortion exists, the correct policy is first to eliminate the distortion through an appropriate tax or subsidy and *then* to apply the instrument indicated by the underlying policy objective. However, this may not always be possible, for the same types of reasons which lead to use of trade restriction where another instrument, such as a direct subsidy, would be more efficient. In such cases, the existence of the distortion may create a stronger presumption in favor of one of the policy alternatives available. For example, if the supply curve overstates social cost, or the demand curve understates social benefit, there is a stronger argument for employing a proportionally distributed quota rather than an equivalent (in achieving the policy objective) tariff/quota.

Similarly, where optimum tariff possibilities are present, the correct policy is first to impose the optimum tariff, and then to use the appropriate form of trade restriction as indicated by the relevant policy objective.

We may summarize the argument as follows: In comparison to a tariff or equivalent quota, a proportionally distributed quota produces an *additional* subsidy to production and a *reduced* tax on consumption. Under competitive conditions, import restriction policies carried out through a proportionally distributed quota rather than a conventional tariff/quota will be more efficient to the extent that the additional production and consumption thus induced tend to contribute to the underlying objective of the policy. If markets are competitive, the import licenses may be given to producers or to consumers without affecting the result, so long as a proportional distribution is employed.

Where the underlying objective of trade restriction is to maintain a desired ratio of

imports to domestic production, a proportionally distributed quota is the optimum instrument to achieve that objective. In addition, the proportionally distributed quota is a *practical* means of implementing the objective, since it can be put into effect without first estimating the supply and demand relations, and does not require revision when changes in these relations occur.[10] However, where a tariff or auctioned quota is replaced by a proportionally distributed quota, the government revenue thus lost must be replaced through another tax.

[10] The cost of pursuing the policy does, however, depend upon the supply and demand relations.

REFERENCES

J. N. Bhagwati, "On the Equivalence of Tariffs and Quotas," in R. E. Baldwin et al., *Trade, Growth and the Balance of Payments—Essays in Honor of Gottfried Haberler*, Chicago 1965, 53–67.

——— **and T. N. Srinivasan**, "Optimal Intervention to Achieve Non-Economic Objectives," *Rev. Econ. Stud.*, Jan. 1969, *36*, 27–38.

H. G. Johnson, "The Cost of Protection and the Scientific Tariff," *J. Polit. Econ.*, Aug. 1960, *68*, 327–45.

———, "The Bladen Plan for Increased Protection of the Canadian Automotive Industry," *Can. J. Econ.*, May 1963, *29*, 212–38.

C. P. Kindleberger, *International Economics*, Homewood 1968.

B. Munk, "The Welfare Costs of Content Protection: The Automotive Industry in Latin America," *J. Polit. Econ.*, Jan./Feb. 1969, *77*, 85–98.

Cabinet Task Force on Oil Import Control, *The Oil Import Question*, Washington 1970.

Effective Protection and Decreasing Returns to Scale

By EITAN BERGLAS AND ASSAF RAZIN*

The concept of effective protection was first developed in the context of a Leontief input-output economy. Only recently has the concept been defined and applied to any constant-returns economy.[1] Jones and Mayer have analyzed effective protection in a general equilibrium model with two commodities. They have shown that effective protection, through its effects on the prices of primary factors, will not necessarily increase the output of the protected industry.[2]

In this paper, we analyze the effects of changes in effective protection in a partial equilibrium context. The partial equilibrium approach to effective protection requires decreasing returns to scale. Otherwise, any increase in effective protection holding primary factor prices constant will result in an indefinite increase in output.[3] The analysis of effective protection in industries with decreasing returns to scale has been carried out mainly for the case of fixed imported-input coefficients.[4] The following analysis does not impose any restriction on the form of the production function. Elsewhere we have analyzed effective protection in a general equilibrium model with more goods than factors. The partial equilibrium analysis to be developed in this paper is preserved unchanged in our general equilibrium model.[5]

Jones' definition of effective protection holds for any constant returns to scale production function. We develop a method by which any decreasing returns production function can be transformed into a constant returns production function by an addition of a fictitious factor of production. Using this transformation, the extension of Jones' definition of effective protection to any decreasing returns production function is straightforward.[6]

In this paper we shall show that in the case of decreasing returns to scale, in a partial equilibrium context, effective protection may *decrease* the output of the protected industry. For this case to arise, there is no need to assume any extreme biases in substitution among factors of production. It is demonstrated that a decrease in production is possible even in the case where the production function is weakly separable between the imported input and primary factors.[7] Since output may decline as a result of effective protection with decreasing returns to scale under fairly general situations, it seems that the case no longer deserves to be regarded as a paradox.

Mordechai Kreinin, James Ramsey, and Jan Kmenta have recently analyzed effective protection in the context of decreasing returns to scale. However, their analysis is

* Department of economics and the Foerder Institute, Tel-Aviv University. We should like to thank Elisha Pazner for helpful comments on an earlier draft. All remaining errors are, of course, our own responsibility. Partial financial support from the Peretz Naftali Fund is acknowledged.

[1] See Ronald Jones. For a similar analysis, see Wolfgang Mayer.

[2] V. K. Ramaswami and T. M. Srinivasan first produced an example of this paradoxical result.

[3] One seemingly exception to this statement is the case in which an industry is producing under constant returns to scale, but some of the factors are specific to the industry and their quantities are fixed. It will become clear in the following analysis that this case is equivalent to the case of decreasing returns to scale analyzed in this paper.

[4] See, for example, Harry Johnson, ch. 14.

[5] See Berglas and Razin. Johnson, ch. 15, has analyzed a similar general equilibrium model, and concluded his analysis with the statement: "It may be that the prevailing mixture of general and partial analysis is the most reasonable approach for practical economists interested in commercial questions" (p. 383). The following analysis is, therefore, in the spirit of Johnson's conclusion.

[6] This method, which is demonstrated to be useful in adapting available analysis for constant returns to scale to decreasing returns to scale in the case of effective protection, may prove to be useful in other cases.

[7] The case of a separable production function is widely used by William Corden in the analysis of effective protection. Within his model Jones shows that even in the separable case protection may lead to a fall in production. We show that the same result may happen in partial equilibrium analysis and provide the necessary condition for this possibility.

restricted to the case of *CES* production function with one imported good and one variable primary factor. In our paper, by introducing the fictitious factor of production, the analysis of effective protection is not restricted to any specific functional form of the production function, nor to one imported intermediate good and one variable primary factor. In their paper, Kreinin et. al. failed to distinguish between changes in value-added and changes in the price of value-added. In our paper, we introduce this distinction; we also show (contrary to their result which is based on a nontraditional definition of effective protection), that changes in effective protection, defined as changes in the price of value-added, are always *proportional* to changes in the rent.

I. Effective Protection for a Decreasing Returns Production Function

In this section, we use Jones' definition of effective protection—which was developed in the context of constant returns to scale—and apply it to a decreasing-returns-to-scale production function.

Let $X=F(K, L, M)$ be a decreasing-returns-to-scale production function of commodity X in its three factors of production: capital (K), labor (L), and an imported intermediate good (M). This production function can then be transformed into a constant-returns-to-scale production function by adding a fictitious factor T as follows:

$$(1) \qquad X = G(K, L, M, T)$$

where G is linear homogeneous in K, L, M, and T. For convenience choose $T=1$. Define G as

$$(2) \quad G(K, L, M, T) = TF\left(\frac{K}{T}, \frac{L}{T}, \frac{M}{T}\right)$$

It can be easily checked that if we multiply K, L, M, and T by $\lambda>0$, output will be multiplied by λ. Since $T=1$, it is obvious that $F(K, L, M)=G(K, L, M, T)$ so that the addition of the factor T does not change the production relation between output X and the factors K, L, and M. Let the prices of X, K, L, and M be denoted by P, R, W, and P_M, respectively. If all prices are given, the output and the quantities of the factors are determined. Denote the resulting rent by P_T. It is obvious that if the industry is faced with the same given prices P, R, W, P_M, and P_T, and the industry is restricted to the use of one unit of T, the same output and quantities of factors will result.

Using this transformation, it is straightforward to apply the definitions of value-added and its price introduced by Jones to a decreasing-returns-to-scale production function. We now define the value-added per unit of output as follows:

$$(3) \qquad a_L W + a_K R + a_T P_T = a_v P_v$$

$$(4) \qquad a_v P_v + a_M P_M = P$$

where a_i is the input-output coefficient of factor $i (i=L, K, T, M)$, a_v is the input-output coefficient of the value-added, and P_v is its price. Following Jones' analysis, the change in P_v (which is the change in effective protection rate) is given by:[8]

$$(5) \qquad \hat{P}_v = \frac{\hat{P} - \theta_M \hat{P}_M}{1 - \theta_M}$$

where $\hat{X}$ denotes dX/X and θ_i is the distributive share of input i.

Substituting equation (3) into (4), differentiating while holding W and R constant, and using the cost minimization property for linear homogeneous production function

$$\theta_L \hat{a}_L + \theta_K \hat{a}_K + \theta_M \hat{a}_M + \theta_T \hat{a}_T = 0$$

we get

$$(6) \qquad \hat{P}_T = (\hat{P} - \theta_M \hat{P}_M)/\theta_T$$

Furthermore, rewriting (6) using equation (5), we get

$$(6') \qquad \hat{P}_T = ((1 - \theta_M)/\theta_T)\hat{P}_v$$

[8] This is equation (A.6′) in Jones. It must be emphasized that this equation is valid for the general case of N primary factors and S imported factors—provided we redefine

$$\theta_M = \sum_{i \in S} \theta_{Mi} \quad \text{and} \quad \hat{P}_M = \sum_{i \in S} \frac{\theta_{Mi}}{\theta_M} \hat{P}_{Mi}$$

where θ_{M_i} and P_{M_i} are the distributive share and the price of imported factor i, respectively. Hereafter θ_M and $\hat{P}_M$ will represent the composite intermediate good.

Thus, the rent always changes in the same direction as effective protection. Clearly,[9] the rate of increase in the rent exceeds the rate of increase of P_v. This result resembles Mayer's conclusion that effective protection has a magnified effect on relative factor rewards.

II. Output Effect of Effective Protection

In order to analyze the output effects, it is necessary to specify in greater detail the characteristics of the fictitious production function.[10] Given fixed W and R, the input-output coefficients are functions of the rent and the price of the imported intermediate input.

$$(7) \quad a_i = a_i(P_T, P_M) \qquad i = K, L, T, M$$

Since $T=1$ we have

$$(8) \quad a_T X = 1$$

Differentiating (8)

$$(9) \quad \hat{X} = -\hat{a}_T(P_T, P_M)$$

Denoting $E_i=(\partial a_T/\partial P_i)(P_i/a_T)$, equation (9) can be rewritten as

$$(10) \quad \hat{X} = -[E_T\hat{P}_T + E_M\hat{P}_M]$$

Substituting equation (6′) into (10) we get

$$(11) \quad \hat{X} = -\left[\frac{(1-\theta_M)}{\theta_T} E_T\hat{P}_v + E_M\hat{P}_M\right]$$

Although E_T and E_M represent the elasticities of the fictitious input-output coefficient a_T, they necessarily have the familiar sign characteristics, namely, $E_T<0$ and $E_M \gtreqless 0$, depending on whether the fictitious factor is a substitute or complement for the imported good.

Equation (11) can be presented alternatively in terms of elasticity of substitution.[11]

$$(11') \quad \hat{X} = -[(1-\theta_M)\sigma_{TT}\hat{P}_v + \theta_M\sigma_{TM}\hat{P}_M]$$

It is straightforward to derive from equation (11) two cases known from existing literature: (a) $E_M=0$; this will be the case if the imported intermediate good has a fixed input-output coefficient. (b) $\hat{P}_M=0$; this is the pure case in which effective protection is a result of changes in tariff on the final product only. In these two cases we have the well-known result that effective protection will increase the output of X. However, this result is not generally correct.

We can now derive the necessary and sufficient conditions for an increase in effective protection to result in a smaller output. From equation (11′) we get a necessary condition $\sigma_{TM}\hat{P}_M>0$. Substituting equation (5) into equation (11′), using $\sigma_{TT}<0$, we get the necessary and sufficient conditions for effective protection to result in $\hat{X}<0$ as follows:

$$(11'') \quad \theta_M\hat{P}_M < \hat{P} < \theta_M\hat{P}_M\left(1-\frac{\sigma_{TM}}{\sigma_{TT}}\right)$$

Given inequality (11″), it is necessary to distinguish among three cases:

Case a: $\hat{P}_M>0$ (therefore $\hat{P}>0$): In the case of $\sigma_{TM}>0$ we have $(1-\sigma_{TM}/\sigma_{TT})>1$, therefore inequality (11″) is nonempty.

Case b: $\hat{P}_M<0$, $\hat{P}\geq 0$: The left-hand side of inequality (11″) is redundant. From the right-hand side of the inequality in the case of $\sigma_{TM}<\sigma_{TT}$, we have $(1-\sigma_{TM}/\sigma_{TT})<0$; therefore, the inequality (11″) is nonempty.

Case c: $\hat{P}_M<0$, $\hat{P}<0$: In the case of $\sigma_{TM}<0$ we have $(1-\sigma_{TM}/\sigma_{TT})<1$, therefore inequality (11″) is nonempty.

We have shown that effective protection in the nonempty cases presented in a, b, and c will result in a decrease in output.

The reader may suspect that paradoxical results occur only in the case of inferior inputs.[12] However, this is not true. Inferiority

[9] Note that $1-\theta_M=\theta_T+\theta_L+\theta_K$

[10] Output effects are hereafter analyzed by the use of a production function. On the level of the industry, the analysis can be done in the following way: (a) If all firms are identical, the whole industry can be described by an aggregate production function. (b) If the industry consists of (identical) constant-returns-to-scale firms, but all use a factor which is specific to the industry, the same analysis applies. (c) If the industry consists of nonidentical firms, then the change in output of the industry is the sum of the changes of output of individual firms.

[11] Observe that $E_T=\theta_T\sigma_{TT}$ and $E_M=\theta_M\sigma_{TM}$, where σ_{ij} is the partial elasticity of substitution of i for j, as defined in R. G. D. Allen, p. 504.

[12] See J. R. Hicks, p. 93.

of M is necessary only in Case b. This can be seen most clearly in the case where $\hat{P}=0$. Then we have a decline of output associated with a reduction of the price of P_M which can occur if and only if M is an inferior factor.[13] A fortiori, M is inferior if $\hat{P}>0$, $\hat{P}_M<0$ and $\hat{X}<0$. In Case a; inferiority of the composite input of K and L is implied when $\hat{P}>\hat{P}_M$. Using zero homogeneity in prices of profit maximizing output, this is equivalent to the case in which $\hat{P}_M=0$, $\hat{W}=\hat{R}<0$ and $\hat{P}>0$. Since $\hat{W}=\hat{R}$, factors K and L can be regarded as a composite factor; thus, by the same reasoning as in Case b above, the composite factor must be inferior.

In Case a, when $\hat{P}_M>\hat{P}$, and in Case c, input inferiority is not required. This can perhaps be seen most clearly in Case a, where the rise in marginal cost will exceed the rise in the price of the final product when effective protection is increased, so that output will decline. It must be emphasized that in these cases "paradoxical" results do not require extreme assumptions about substitution between factors of production. We illustrate this result for a decreasing-returns Cobb-Douglas production function.[14]

III. The "Paradox" with a Separable Production Function: An Illustration

Let $X=F(K, L, M)=K^{\alpha_K}L^{\alpha_L}M^{\alpha_M}$ where $\sum_i \alpha_i<1$ and $\alpha_i>0$ $(i=K, L, M)$. Then

$$G(K, L, M, T) = K^{\alpha_K}L^{\alpha_L}M^{\alpha_M}T^{\alpha_T}$$

where $\alpha_T=1-\sum_i \alpha_i$. Under competition, the rent is equal to the marginal productivity of the fictitious factor T:

$$\alpha_T \frac{P}{a_T} = P_T \tag{12}$$

Substituting equations (6) and (12) into equation (8) and differentiating, we get:[15]

$$\hat{X} = \frac{(1-\alpha_T)}{\alpha_T}\hat{P} - \frac{\alpha_M}{\alpha_T}\hat{P}_M \tag{13}$$

The necessary and sufficient condition for $\hat{X}<0$ and $\hat{P}_v>0$ is

$$\frac{1}{\alpha_M} > \frac{\hat{P}_M}{\hat{P}} > \frac{1-\alpha_T}{\alpha_M}$$

$$\text{and} \quad \hat{P}_M > 0$$

Observe that since $\alpha_M<1-\alpha_T$, it is necessary in this case that $\hat{P}_M>\hat{P}$.

IV. Concluding Remarks

We have shown that the standard measure of effective protection is a good predictor of changes in the rent of the decreasing returns industry. However, changes in output are not uniquely related to changes in this measure.[16] Furthermore, increase in effective protection in many cases will lead to a decline in output. Ruling out the possibility of inferior inputs, production will expand whenever the rate of increase in output price is greater than or equal to the rate of increase of imported input price.

REFERENCES

R. G. D. Allen, *Mathematical Analysis for Economists*, London 1938.

E. Berglas and A. Razin, "A Trade Model with Variable Returns to Scale," *Econ. Rec.*, Mar. 1973, *49*, 126–34.

W. M. Corden, *The Theory of Protection*, London 1971.

J. R. Hicks, *Value and Capital*, 2d ed., London 1946.

H. G. Johnson, *Aspects of the Theory of Tariffs*, London 1971.

R. W. Jones, "Effective Protection and Substitution," *J. Int. Econ.*, Feb. 1971, *1*, 59–81.

[13] See Moses Syrquin, proposition 1.

[14] The example of the Cobb-Douglas production function, which is weakly separable between M and other inputs, demonstrates that in the case of decreasing returns separability does not insure that effective protection will increase output. It may be recalled that separability was introduced as a sufficient condition for effective protection to result in increased output in the case of constant returns to scale.

[15] The same result can be obtained from equations (5) and (11′) using $\theta_i=\alpha_i(i=K, L, M, T)$ and $\sigma_{TM}=1$ and $\sigma_{TT}=-(1-\alpha_T)/\alpha_T$ for the Cobb-Douglas case. Observe that equation (13) is identical to equation (8) in Kreinin et. al.

[16] This result is contrary to the claim by Kreinin et. al.: "It is changes in the rate of profit that govern adjustments in domestic output" (p. 891).

M. E. Kreinin, J. B. Ramsey, and J. Kmenta, "Factor Substitution and Effective Protection Reconsidered," *Amer. Econ. Rev.*, Dec. 1971, *61*, 891–900.

W. Mayer, "Effective Tariff Protection in a Simple General Equilibrium Model," *Economica*, Aug. 1971, *38*, 253–67.

V. K. Ramaswami and T. M. Srinivasan, "Tariff Structure and Resource Allocation in the Presence of Factor Substitution," in J. Bhagwati et al., eds., *Trade, Payments and Welfare*, Amsterdam 1971.

M. Syrquin, "A Note of Inferior Inputs," *Rev. Econ. Stud.*, Oct. 1970, *37*, 591–93.

External Diseconomies in Competitive Supply: Comment

By George A. Hay and John J. McGowan*

In a recent article in this *Review*, Charles Goetz and James Buchanan (G-B) assert that " . . . the standard description of misallocation in the presence of external production diseconomies is misleading . . . " because these externalities produce a " . . . combination of exchange-inefficiency with production-inefficiency [which] renders the construction of correction devices much more difficult" (p. 889). Stated otherwise their contention is that with external diseconomies that are internal to an industry, i.e., those that each firm in an industry inflicts on other firms in the same industry, a competitive regime in the presence of a per unit tax on output designed to eliminate the difference between private and social marginal cost will not achieve a Pareto optimum. The competitive equilibrium even after the imposition of the corrective tax lies inside the production possibilities frontier.[1]

The purpose of this note is to suggest that the G-B analysis while technically correct is based on a set of assumptions that differs in a particular aspect from what we believe to be the assumptions of the neoclassical paradigm. Moreover within the framework of the neoclassical model we show the standard prescription regarding a corrective tax to be correct. As it turns out, the difference between the G-B model and the neoclassical model is likely to be negligible in any meaningful application but in any case we show that in the G-B model, a lump sum tax imposed on top of the standard Pigovian per-unit tax will restore full Pareto optimality.

G-B assume that the total cost function of any firm has the following specific functional form:

$$(1)\quad c_i = k + q_i^a + Q_i^b q_i \qquad a > 1,\ b > 0$$

where q_i is the firm's own output, Q_i is the aggregate output of all *other* firms in the industry, and where $b>0$ indicates the presence of a technological external diseconomy. All firms are assumed to have the same cost function.

A general version of such a cost function is the following:

$$(2)\qquad c_i = g(T/n, T)$$

where c_i is total costs for firm i, T is total industry output and n is the number of firms in the industry. Since all firms are assumed to have the same cost function, efficiency requires that output be distributed evenly among them. Thus, with n firms, each will produce T/n units.[2]

Dropping the subscript i, we can write total industry costs for any output T as:

$$(3)\qquad TC = nc = ng(T/n, T)$$

The first condition for Pareto optimality is that of production efficiency, i.e., any output T should be produced at minimal

* Associate professors, Yale University. We are grateful to William Brainard, Joseph Stiglitz, and Martin Weitzman for helpful comments.

[1] As G-B point out, the issue involves technical rather than pecuniary externalities since it is well known that the latter cause neither exchange nor production inefficiency. Moreover we assume G-B intended to restrict their analysis to external diseconomies in the production of a final (consumption) good rather than an intermediate good. We think it is clear that the neoclassical writers have recognized that external diseconomies in an intermediate goods industry would cause the unregulated competitive equilibrium to lie inside the economy's production possibilities frontier, since producers of final products would be led to choose inefficient input combinations. However even in this case the neoclassical writers, correctly we believe, would claim that a simple Pigovian tax on the intermediate product in question would restore full Pareto optimality.

[2] Near the end of their paper, G-B generalize their original cost function in the form: $f(T/n,\ T-T/n)$. Obviously, without further restrictions on g or f, the two are indistinguishable. We use (2) because it highlights the key neoclassical assumption more effectively, but the identical result can be obtained using the G-B version.

total cost to society. The policy variable for the "social planner" is T/n, or since T is given, n. Thus we wish to minimize total industry costs with respect to n for a given T. This requires:

$$(4)\quad \frac{\partial TC}{\partial n} = c + n \cdot \left(\frac{\partial c}{\partial T/n}\cdot\frac{\partial T/n}{\partial n} + \frac{\partial c}{\partial T}\cdot\frac{\partial T}{\partial n}\right)$$
$$= c + ng_1(-T/n^2)$$
$$= c - g_1 \cdot T/n = 0$$

where $g_1 = \dfrac{\partial c}{\partial T/n}$

$$(5)\quad \therefore\ g_1 = \frac{c}{T/n}$$

but $c/(T/n)$ is average costs, AC, for each firm. Therefore, efficiency in production requires that production take place in each firm up to the point where $g_1 = AC$.

The second condition for Pareto optimality is efficiency in exchange. This in turn requires that total industry output be expanded to the point where price is equal to social marginal cost. The latter is defined by:

$$(6)\quad \frac{\partial TC}{\partial T} = n\left(\frac{\partial c}{\partial T/n}\cdot\frac{\partial T/n}{\partial T} + \frac{\partial c}{\partial T}\right)$$
$$= n\left(\frac{g_1}{n} + g_2\right)$$
$$= g_1 + ng_2$$

where $g_2 = \dfrac{\partial c}{\partial T}$

Thus the condition for exchange efficiency reduces to:

$$(7)\quad P = g_1 + ng_2$$

On the other hand, profit maximization for the firm requires that production be expanded to the point where price equals *private* marginal cost, MC. The latter is defined by:

$$(8)\quad MC = \frac{\partial c}{\partial q_i} = \frac{\partial c}{\partial T/n} + \frac{\partial c}{\partial T}\cdot\frac{\partial T}{\partial q_i}$$
$$= g_1 + g_2\frac{\partial T}{\partial q_i}$$
$$= g_1 + g_2 \left(\text{since } \frac{\partial T}{\partial q_i} = 1\right)$$

where we have used the q_i term to indicate that the marginal cost for any one firm producing T/n is defined in terms of increasing only its own output, with all other firms continuing to produce at T/n.[3]

We assert that a key assumption of the neoclassical paradigm is that firms regard any increase in costs caused by the diseconomy effect of higher industry output as an exogenous *shift* in their cost curves, unrelated to changes in their own or industry output.[4] This is equivalent to the assumption that firms perceive g_2 as equal to zero.[5]

If this assumption about firms' perception of g_2 is correct, the first condition for Pareto optimality will be fulfilled, since competitive equilibrium will be established where $P = MC = AC$, which reduces to $g_1 = c/(T/n)$ as in (5). (Of course if g_2 is negligibly small, the assumption is unimportant since we see from (8) that equilibrium output and the Pareto optimal output differ by the order of g_2. Moreover as we shall see, there is strong reason to believe that g_2 will be negligibly small in meaningful applications.)

To secure exchange efficiency, the neoclassical economists prescribed a per-unit tax on output. If that tax is set equal to ng_2,

[3] Implicit in (8) is the assumption that the diseconomy effect is a function of total industry output independent of its distribution among firms. This assumption is not critical to the main results. We could have written (1) as $c_i = g(q_i, q_{i+1}, q_{i+2}, \ldots, q_{i+n-1})$. The first-order conditions would be more complex but their nature would not change. The G-B cost function does *not* assume that the diseconomy is independent of the distribution of total output but the nature of the required tax is unaffected by this asymmetry.

[4] John Chipman in an elegant analysis of the neoclassical argument makes this assumption most explicit.

[5] The spirit of this assumption is identical to that under which a firm in perfect competition assumes that expansion of its own output will not affect market price.

we have as a condition for private profit maximization:

$$(9) \quad P - ng_2 = g_1$$

$$\therefore \; P = g_1 + ng_2 \text{ as required by (7)}$$

Thus, under what we believe to be the assumptions of the neoclassical paradigm for external diseconomies, the prescription of a per-unit tax equal to the difference between private marginal cost and social marginal cost is the correct one and is sufficient to restore the economy to a Pareto optimal position.

We note now that where n is large, as we would expect in perfect competition, ng_2 can be significant even though g_2 is negligible. Moreover, *if g_2 is not negligible*, a tax equal to ng_2 would be so large as to probably require a complete shutdown of the industry, or at best an output so small that atomistic competition would not be feasible and a totally different set of controls would be required. Therefore, in any case in which the prescription of a Pigovian tax could be a feasible remedy, there is reason to believe that g_2 would in fact be negligibly small, and the neoclassical assumption would not be unrealistic.

Despite the above, if firms do *not* ignore g_2, then each firm expands production to the point where $P = g_1 + g_2$, and the tax required to restore exchange efficiency is changed to $(n-1)g_2$. In that event, the production efficiency condition will *not* be satisfied since then we will have:

$$(10) \quad P - (n-1)g_2 = g_1 + g_2 = \frac{c}{T/n}$$

thus:

$$(11) \qquad g_1 = \frac{c}{T/n} - g_2$$

violating (5).

This situation can be remedied, however, if we impose in addition to the per-unit tax, a lump sum tax equal to $g_2 \cdot T/n$. Doing so will cause competitive equilibrium to be established where:

$$(12) \; P - (n-1)g_2 = g_1 + g_2 = \frac{c + g_2 \cdot T/n}{T/n}$$

resulting in $g_1 = c/(T/n)$ as required by (5).

Thus if firms regard the level of diseconomy as beyond their control, which we believe is the assumption of the neoclassical paradigm, the Pigovian tax guarantees Pareto optimality. Where firms recognize the relation between industry output and their own costs, the competitive equilibrium with the Pigovian tax lies inside the production possibilities frontier by a factor of the order of g_2 which it seems reasonable to expect will be negligibly small when the number of firms is large. In any case, full Pareto optimality can be restored by a lump sum tax on top of the Pigovian per-unit tax.

Examination of the G-B model reveals that the *only important difference* between it and the neoclassical model is precisely the assumption about g_2. In the G-B model the diseconomy effect of an increase in a firm's output (and therefore total industry output) on its own costs is blended into the cost function along with direct costs and both are assumed to be fully recognized by the firm in deciding how far to expand output.

Thus G-B are correct in asserting that *in their model* the competitive equilibrium, even with the Pigovian tax equal to the difference between private marginal cost and social marginal cost, lies inside the production possibilities curve. What we have shown is that the remaining inefficiency should be negligible. In any case we have also shown that a lump sum tax added to the per-unit tax will be sufficient to restore full Pareto optimality.

REFERENCES

J. S. Chipman, "External Economies of Scale and Competitive Equilibrium," *Quart J. Econ.*, Aug. 1970, *84*, 347–85.

C. J. Goetz and J. M. Buchanan, "External Diseconomies in Competitive Supply," *Amer. Econ. Rev.*, Dec. 1971, *61*, 883–90.

External Diseconomies in Competitive Supply: Comment

By ALAN NICHOLS*

Recently in this *Review*, Charles Goetz and James Buchanan advanced the proposition that output-generated external diseconomies may be associated with a production possibilities curve internal to an attainable curve and that under competition it would be the former on which equilibrium would occur. This in turn they take to imply that the conventional tax-bounty analysis offers, in general, an incorrect remedy in that it applies to the wrong production possibilities curve.

The present discussion does not deny that the results indicated by Goetz and Buchanan occur, as they too note, in the context of *input*-generated external diseconomies nor that their results may be conceivable in the context of output generated external diseconomies.[1] We do however deny that the analytical structure they offer in support of their conclusion is admissible.

The analytical case Goetz and Buchanan make is, in terms of Dean Worcester's classification, that of "Externalities which are internal to a specific industry" (p. 884). They employ the crucial specification, on which their results turn, that the costs of each firm (where firms are identical) depend on a function which includes the output of *other* firms, as distinct from total industry output, as an argument. Where c_i, q_i, Q_i are, respectively, firm total cost, firm total output, and output of all other firms corresponding to any particular firm, i, we have

$$c_i = f(q_i, Q_i) \tag{1}$$

For full competitive equilibrium the $mc = ac$ condition requires

$$c_i/q_i = \partial f/\partial q_i \tag{2}$$

Holding industry output, T, constant we have the number of firms, n, as

$$n = T/q_i \tag{3}$$

and

$$Q = (T - q_i) \tag{4}$$

Efficient production, subject to the industry output constraint *and* the particular specification that Q is output only of other firms, requires

$$c_i/q_i = \partial f/\partial q_i - \partial f/\partial Q \tag{5}$$

Thus the output of each firm should exceed that attained under competition or, in other words, for given industry output there should be fewer firms and each should be larger. From this it follows that, since firms are too small, the industry is operating below an attainable production possibilities curve; hence the Goetz and Buchanan conclusions.

To evaluate the Goetz and Buchanan analysis it is necessary first to note that a minuscule change in specification eliminates their conclusions. Thus rewrite (1) as

$$c_i = f(q_i, Q + q_i) \tag{6}$$

and the constraint that T is a constant now produces conventional results, not those they identify. With total industry output the source of external diseconomies we no longer have the curious economies of scale that are implicit in their specification. Since it was the failure to realize these scale economies that presented the interior production possibilities curve in the first place, it is, in turn, not surprising that they find that "... any attempt to impose a per unit corrective tax on the output of firms in the industry will fail" (p. 888). Here their problem is not that of external output diseconomies; it is that the firms are too small. Meanwhile the fact that there are external output diseconomies is a priori neither here nor there for firm size, nor is firm size a priori relevant to external output diseconomies. It is only the "other firm" specification that makes it so.

One might ask is there any advantage in

* Professor of economics, Central Michigan University.

[1] To say that they may be conceivable is not the same as saying that we can conceive of them. See reference in fn. 2, below.

associating external output diseconomies internal to an industry only with output of *other* firms, apart from the advantage of being totally literal? The literature, to our knowledge, discloses no example.[2] The common oil pool case clearly includes the firm's own output summed with the output of other firms. But our failure to conceive of an example need hardly be convincing of the unreality of the case. What may be more convincing are the very curious and quite specific implications of the other firms analysis. As noted, the Goetz and Buchanan firm economies are not ordinary economies of scale. They are not present because a firm is larger. They are not even present because a firm is larger relative to other firms singly considered. They only become present because a firm is larger relative to all other firms together. Formally, where industry output is a variable and where we continue to treat all firms as identical we note that these "economies of scale" would occur whenever the percentage increase in output per firm is less than the percentage decrease in the number of firms. Diseconomies of scale would occur when the inequality is reversed. We find it hard to believe that there is even any point in trying to fill such an economic box.[3]

What is to be said about the introduction of a construct which is unnecessary in explaining the phenomenon with which it deals; which has bizarre and implausible implications; which creates havoc in the analysis of widely accepted policy remedies; and, finally, which is associated with no support other than that which comes from the fact that it has been algebraically specified? We submit that such a construct should be rejected out of hand. What is novel in Goetz and Buchanan, unfortunately, goes with it.

REFERENCES

M. Blaug, *Economic Theory in Retrospect*, Homewood 1968.

C. J. Goetz and J. M. Buchanan, "External Diseconomies in Competitive Supply," *Amer. Econ. Rev.*, Dec. 1971, *61*, 883–90.

D. A. Worcester, Jr., "Pecuniary and Technological Externality, Factor Rents and Social Costs," *Amer. Econ. Rev.*, Dec. 1969, *59*, 873–85.

[2] A very convenient survey is in Mark Blaug's discussion.

[3] If Goetz and Buchanan simply wanted to say that, *entirely apart* from output generated external diseconomies, it is possible that firm size might be suboptimal under competition, we would concede that this could be the case and that increases in output need not provide a remedy. But then they would have been saying nothing about output generated external diseconomies, the primary subject of their enquiry. Instead they have slipped in, without comment, a particular algebraic artifice which *definitionally* links firm size and output generated external diseconomies; they do not trouble us with either an empirical or an intuitive linkage. All of this despite the fact the tiniest of changes in their definition eliminates their results altogether.

External Diseconomies in Competitive Supply: Comment

By A. Ross Shepherd*

In a recent article in this *Review*, Charles Goetz and James Buchanan argued that "... familiar descriptions of the nonoptimality properties of competition in the presence of technological external diseconomies are misleading in that they focus exclusively on exchange inefficiency and neglect production inefficiency" (p. 883). With respect to production inefficiency caused by output-generated diseconomies they conclude: "Any fully corrective tax must be more complex than has been realized, since not only marginal cost but also the firm's perception of the *shape* of its average cost curve must be modified" (p. 888). The purposes of this comment are: 1) to reveal the implicit assumption on which this result of Goetz and Buchanan is based; and 2) to show that a more appealing assumption implies the absence under perfect competition of production inefficiency due to output-generated diseconomies.

Goetz and Buchanan (hereafter, G-B) assume an economy comprised of competitive industries. In one industry the (identical) constitutent firms reciprocally impose output-generated technological external diseconomies that are assumed to be "... contained within a well-defined industry category" (p. 883).[1] For each of these firms the total cost function is written (p. 888):

$$c_i = f(q_i, Q_i) \tag{1}$$

where c_i is the total cost of the ith firm, expressed in units of homogeneous input; q_i is the output of the ith firm; and Q_i, the diseconomy argument $(\partial f/\partial Q_i > 0)$, is the total output of all other firms in the industry. According to G-B long-run competitive equilibrium obtains with (p. 888):

$$\frac{c_i}{q_i} = \frac{\partial f}{\partial q_i} \tag{2}$$

That is, with each firm's *ceteris paribus* long-run average cost equal to its *ceteris paribus* long-run marginal cost. Evidently G-B suppose that the competitive firm behaves as if it believed that its output decisions left other firms' outputs unaltered. But this requires that each of the firms ignores repeated observations to the contrary, which is a scientifically unappealing assumption.

The basic reasoning to be applied here was first worked out many years ago by Joan Robinson. The *sine qua non* for a perfectly competitive industry is that each constituent firm faces a demand curve of sufficiently great elasticity that it differs negligibly from a horizontal demand curve. It follows that if the continuous, negatively sloped market demand curve facing the competitive industry is to be taken seriously, as it is in theoretical formulations, it must be assumed that market output is negligibly affected by any one firm's output decisions. And this condition is fulfilled precisely to the extent that an autonomous change in the output of one firm automatically elicits a countervailing *change in the output of the other firms in the industry.*

For any given market demand curve this countervailing reaction is increased as the number of firms increases and the slope of the firms' marginal cost curves decreases. In addition to the ith firm let there be n (identical) firms in the industry. Then:

$$Q_i = \sum_{j=1}^{n} q_j = nq_j \tag{3}$$

If the ith firm changes output, market price (P) changes inversely and the induced change in output by the other n firms will be:

$$\frac{dQ_i}{dP} = n\frac{dq_j}{dP} \tag{4}$$

* Professor of economics, University of Missouri–Kansas City.

[1] This assumption eliminates the possibility of pure exchange inefficiency because all diseconomies (social costs) are internal to the industry and therefore do not distort commodity price ratios away from ratios of social costs.

Evidently the absolute value of this product will increase with the number of firms (n) and the reciprocal of the slope of the marginal cost curves (dq_j/dP). Alternatively put, the greater the number of firms, and the less steep are their marginal cost curves, the greater is the induced countervailing change in output and the less will market price be affected by an autonomous change in output by the ith firm.[2] The perfectly competitive firm, precisely to the extent that it is perfectly competitive, will affect the output of other firms in the industry. Unless we assume that entrepreneurs are obtuse, we must suppose that they will become aware of this reaction, in which case the competitive equilibrium described by G-B is invalid.[3]

If we restate the G-B model without entrepreneurial obtuseness we note:

$$Q_i = Q_i[P(q_i)] \tag{5}$$

Therefore:

$$\frac{dc_i}{dq_i} = \frac{\partial f}{\partial q_i} + \frac{\partial f}{\partial Q_i}\left(\frac{dQ_i}{dP}\,\frac{dP}{dq_i}\right) \tag{6}$$

As the competitive industry approaches perfection the bracketed expression on the right-hand side of (6) approaches -1 and (6) approaches:

$$\frac{dc_i}{dq_i} = \frac{\partial f}{\partial q_i} - \frac{\partial f}{\partial Q_i} \tag{6a}$$

[2] It is perplexing that textbooks fail to explain this fundamental point. The usual textbook formulation should leave readers asking why an increase in output by the competitive firm does not affect price perceptibly whereas that same increment in the same market would supposedly affect price perceptibly if provided by a monopolist.

[3] The competitive equilibrium described by G-B, shown at E^{**} in their Figure 2, p. 886, and at e_i in their Figure 3, p. 887, is unstable. Suppose that all entrepreneurs are obtuse and therefore fail to recognize the connection between the output of one firm and the output of other firms in the industry. Any unplanned change in output of any firm will disturb the competitive equilibrium and market forces will not be generated to restore it. Instead, the market reactions discussed in the text above will work to move the firm along $\bar{ac}$ in G-B's Figure 3.

In the limiting case, then, competitive equilibrium obtains with:

$$\frac{c_i}{q_i} = \frac{\partial f}{\partial q_i} - \frac{\partial f}{\partial Q_i} \tag{7}$$

for each firm, which is the social (production) optimum described by G-B, p. 888.[4] Speaking in terms of G-B's Figure 3, p. 887, each firm located at e_i will see what appears to be the opportunity for profitable expansion. As all firms expand simultaneously market price falls and market quantity demanded increases. The expanding firms reciprocally inflict additional diseconomies, which shift the $\overline{ac}$ and $\overline{mc}$ curves upward. Competitive equilibrium obtains with the now larger firms each producing at minimum unit cost ($\bar{e}'$) with $e_i > \bar{e}' > \bar{e}$. In this situation there is no socially cheaper way to produce the current industry output. Production inefficiency caused by an output-generated external diseconomy is automatically eliminated by the perfectly competitive market.

In sum, if the externality is output-generated economists can continue to use the familiar descriptions of the nonoptimality properties of competition in the presence of technological external diseconomies without being misled, and they can continue to recommend the standard Pigovian corrective without misleading others.[5]

REFERENCES

C. J. Goetz and J. M. Buchanan, "External Diseconomies in Competitive Supply," *Amer. Econ. Rev.*, Dec. 1971, *61*, 883–90.

J. Robinson, "What Is Perfect Competition?" *Quart. J. Econ.*, Nov. 1934, *49*, 104–20.

A. R. Shepherd, "Corrective Taxes Under Monopoly," *Quart. Rev. Econ. Bus.*, summer 1972, *12*, 33–38.

[4] In G-B's equation (12), p. 888, $\partial f/\partial Q_i$ appears twice as $\partial f/\partial Q$. From the context it is clear that this is a typographical error.

[5] Provided, of course, that perfect competition prevails. For a recent discussion of corrective taxes in the absence of perfect competition, see Shepherd.

External Diseconomies in Competitive Supply: Reply

By CHARLES J. GOETZ AND JAMES M. BUCHANAN*

The many comments stimulated by our recent paper call into question its concluding assertion that "no real paradigm has existed in this area of the theory of competitive supply." Several critics allege that our conclusions as to the existence of "production inefficiencies as well as exchange inefficiencies" depend on a particular form of the firm's cost function. An alternative form, said to be more in keeping with the standard assumptions of competitive theory, is shown to restore the orthodox analysis in its pristine purity and simplicity, including its implications for the efficacy of Pigovian corrective taxes. While we concede that the critics' specification of the cost function does rescue the orthodox analysis, we dispute whether their formulation more nearly reflects the traditional behavioral assumptions of competitive theory. Indeed, while legitimate argument on the point may exist, we continue to feel that our own formulation embodies the traditional assumptions. In any event, we are unconvinced that the alternative premise should be favored simply because it restores the neatness of the Pigovian prescriptions. The real issue is which of the formulations more accurately captures the behavioral rule followed by competitive decision makers.

I. What is Parametric in Competitive Theory?

We and our critics can probably agree that the essential notion of competition is that all decision makers are "price takers," regarding themselves as unable to affect price by their own quantity adjustments. An alternative expression of this notion is that competitive behavior is *parametric*, taking certain variables as behavioral constants which are part of an exogenously fixed "state of nature." The dispute arises over precisely what variables *in addition to price* are to be included in the set of fixed parameters. We, for instance, include the production of other firms, Q_i, as part of the state of nature faced by the decentralized decision maker, firm i. The opposing school is troubled by one of the logical *lacunae* of competitive theory which arises when we express price as a function of total supply T

$$p = D(T) = D(Q_i + q_i)$$

where the ith firm's own product q_i is allowed to influence T and $\partial p/\partial q_i$ cannot be taken as mathematically equal to zero.

One seemingly appealing way out of the mathematical bind is then to assume, as do our critics, that $\partial T/\partial q_i$ must be zero, that external production adjusts to the ith firm's production q_i in a counterbalancing manner:

$$Q_i = T - q_i$$

In production theory, there is even an apparently plausible rationalization of this assumption since such adjustment may indeed be expected to occur 1) in the long run, and 2) *provided* that the industry also faces constant factor costs.[1] However, the imposition of that latter proviso should suggest to the strict mathematical logician that the price-constancy paradox has merely been swept further under the rug, into input markets.

These difficulties in competitive theory are rendered more emphatic by noting that the analogous problem exists in demand theory. Here, one considers the impact of a single competitive buyer on a good whose supply can, for simplicity, be assumed to be fixed. In the demand model, there is certainly no rationale at all for assuming that other consumers will make exactly counterbalancing adjustments in their bids in order to preserve

* Virginia Polytechnic Institute and State University.

[1] On the assumption of U-shaped average cost curves, a change in the distribution of production among firms will affect the total input requirements for a given output. Changes in factor prices will then be transmitted to firms' cost curve. In general, the original industry output equilibrium would then be disturbed.

price constancy as a single buyer alters his own bids. Mathematically, the single buyer does have an impact on price.

The point of this discussion is that some economists regard the fixity of aggregate supply assumption as a necessary implicit corollary of price-taking behavior. However, as we view it, this fixity assumption is not only not necessary, but does not even fully satisfy the apparent logical demands out of which it purports inexorably to emerge. If the insoluble logical paradoxes presented in the strict mathematical model are regarded in less purist fashion, our interpretation of the competitive assumption as a behavioral rule-of-thumb offers considerable appeal. On this assumption, the competitive decision maker controls only his own personal activity variables and regards all other variables, including output of competitors, as part of a state of nature over which he exercises no control.[2] Our so-called *ceteris paribus* cost functions reflect this conceptualization of competitive behavior, one which is certainly a defensible characterization of the neoclassical competitive model and which we find more descriptive than the suggested alternative.

II. Related Issues

Quite apart from the question of what should be parametric in competitive cost functions, some critics feel that our formulation does not correspond to the traditional notion of industry externalities as being dependent on total industry output rather than on merely the "external" element consisting of other firm's output. There is really no problem here if it is recognized that our assumption about the *single-firm* production function does imply an aggregate *industry* production function in which external costs are determined by total industry output. However, our particular formulation was also specifically aimed at raising an additional question as to whether the behavioral situation of the competitive decision maker does not render the distribution of output among firms inefficient. The neoclassical price theory literature was apparently not concerned with this question and did not precisely define the point at issue here. Hence it would be difficult to adduce evidence from the literature as to which of the alternative formulations was intended. For instance, we fail to see any necessary incompatability between our own model and the standard examples such as the oil pool case.[3] The modern reader should find this ambiguity in the literature unsurprising and judge the alternative conceptions on their own merits.

Our own position would, in any event, attach relatively little empirical importance to what we term the strict "output-generated" externalities customarily discussed under the term "industry diseconomies of scale." Quite clearly, most of the real world examples are input-generated and the simplistic Pigovian output tax is of little practical interest. Our original theoretical argument was developed in the course of a methodological exploration of whether the *average* cost conditions of competitive equilibrium pose any problems within an externalities model. The output-generated case was simply the most straightforward model in which to argue that they do.

Nevertheless, the present argument about the proper specification of behavioral parameters also has implications for the input-generated case. Although we discussed the input-generated case briefly, a more extensive development of essentially the same point was published almost simultaneously by Laurence Schall. Schall's piece does not consider average cost conditions explicitly and leaves the impression that the production inefficiency effect of an externality-generating input may be absent when all firms are identical and the marginal effects are,

[2] Our position seems to be supported by Lester Telser who rejects the price-output constancy definition of competition: "A better definition is that firms act independently. This is the notion given precision by the Cournot-Nash model" (p. 276).

[3] In this case, our model predicts two separate effects. Too much pumping effort will be undertaken and too much oil pumped out of the common pool. In addition, however, there will also tend to be too many separate wells drilled. For any given amount of oil pumped, costs per barrel could be reduced by reducing the number of holes.

consequently, mutually offsetting. Albeit the identical firms assumption is itself a special case, discussions of the competitive model are frequently framed in these terms. What, then, do average cost conditions suggest as a qualification of Schall's conclusion?

If, as our critics allege, competitive firms should properly be conceived of as regarding total industry output as a parameter, then Schall's conclusion requires no qualification. However, if the input-generated case is reformulated analogously to our output-generated case, it can be shown that even identical firms do not achieve production efficiency under the postulated conditions. The nature of the inefficiency cannot, of course, be stated in terms of firm size in this case. Hence, the basic methodological controversy under discussion here does generalize to what may be regarded as more plausible cases of production externalities.

This may also be an appropriate opportunity to make two points which have been the subject of some misunderstanding. Nothing in our analysis suggested that a Pigovian tax could not be used to *improve* resource allocation. Also, although our Figures 2–3 were generated from an explicit production function, and therefore represent conceivable results, we make no claims for the empirical magnitude of the inefficiency effect demonstrated in our theoretical model.

The comment by George Hay and John McGowan does raise a further theoretical issue which touches upon the *magnitude* of the inefficiency implied when the Goetz-Buchanan formulation is employed. This inefficiency depends on the firm's perception of the extent to which its own costs are influenced by changes in industry output, the relevant derivative being denoted as g_2 in the Hay-McGowan terminology. They first argue that g_2 must be negligably small because, otherwise, the appropriate n firm corrective tax of ng_2 would either require a complete shutdown of the industry or permit a number of firms so small that the competitive assumptions would be violated. Then, they point out that the difference between the efficient single-firm solution and the inefficient one varies with the order of g_2. The implication is that the inefficiency noted in our original price is not, in any event, anything to worry about. (A somewhat bizarre result to emerge from defenders of Pigovian orthodoxy.) It seems to us that, on this point, our critics are confused between the *absolute value* of an excess social cost and the *ratio* of the excess social cost to efficient social cost.

Consider a simple example. There are 1000 firms producing whizbangs under conditions of output-generated external diseconomies. Suppose we know that the incremental costs analogous to g_2 for a one-unit increase in industry output T are 1¢. Then the corrective tax is \$10 per unit. In order to believe that g_2 "must" be small, one should be willing to believe that any such industry which pays taxes of \$10 per unit, is either (a) eliminated or (b) reduced to a very small number of firms. Clearly, this argument is not a plausible generalization. As to the percentage cost inefficiency generated, one may well suspect, based on the 1¢ figure, that it is small. However, when industry output T is large, the absolute magnitude of the excess resource cost may nonetheless be very substantial because a small excess cost per unit must be multiplied by the appropriate output figure. In sum, we do not agree that theoretical considerations permit one to assert a priori that the Goetz-Buchanan type of production inefficiency is negligibly small, even though we are similarly unwilling to make the converse assertion that the impact is empirically large.

III. Concluding Remarks

Science advances by attacks on its orthodoxy and there is an important role to be played by the conservative defenders of the faith. If it is not too gross a misplaced metaphor, we would nevertheless suggest that the selection of a behavioral premise on the basis of whether it justifies widely held analytical conclusions is a case of letting the tail wag the dog. Which analysis should be labelled "inadmissible" or "an empty box" depends on how one resolves an ambiguous behavioral aspect of the competitive model. Our concep-

Bond Price Volatility and Term to Maturity: A Generalized Respecification

By MICHAEL H. HOPEWELL AND GEORGE G. KAUFMAN*

One of the more common and accepted generalizations in the mathematics of bond prices is that "... for a given change in yields, the fluctuations in market price will be greater the longer the term to maturity" (W. C. Freund, p. 51). Similar statements may be found in almost every book discussing the mechanics of bond prices.[1] This paper demonstrates that this proposition does not hold in all cases, and we derive the general relationship.

In his important book on the term structure, Burton Malkiel specifies the above relationship between bond prices and maturity as the second of his five general theorems:

> For a given change in yield from the nominal yield, changes in bond prices are greater, the longer the term to maturity. [p. 54]

However, shortly afterwards he notes:

> This theorem is not generally true when yield changes are measured from a base other than the nominal yield. In particular, when bonds are selling at a discount ... it is possible to find cases where longer-term securities are actually less sensitive to a given change in market interest rates than are shorter issues. [p. 55]

Malkiel does not consider this point to have much "practical importance" and does not explore the reasons for its existence. Moreover, it does not prevent him from proceeding to this third theorem:

> The percentage price changes described in Theorem 2 increase at a diminishing rate as N (maturity) increases. [p. 55]

However, the high market rates of interest in recent years have given greater practical importance to the inverse relationship between term to maturity and change in bond price. In a recent book directed at portfolio managers, Sidney Homer and Martin Leibowitz demonstrate this phenomenon numerically for hypothetical bonds. They show that when market interest yields are reduced by 25 percent from 9.5 percent to 7.1 percent, "... 20-year 1's, 2's, 3's, and 4's (coupons) are more volatile than 30-year 9's" (p. 52). The authors find no easy answer for this "puzzling" behavior and conclude that "... the volatility of any conventional high-grade bond results from the interaction of three factors: maturity, coupon, and the starting level of yields" (p. 51).

The apparent inability of analysts to explain this unusual bond price pattern reflects an incomplete understanding of the mathematics of bond prices. While price volatility is related to the time structure of a bond, it is not mathematically related to term to maturity in any simple way. Rather, it is proportionately related to the duration of the bond.[2] The general theorem may be stated:

> *For a given basis point change in market yield, percentage changes in bond prices vary proportionately with duration and are greater, the greater the duration of the bond.*

The validity of this statement for all bonds

* University of Oregon. We are indebted to our colleague, G. O. Bierwag, for helpful comments throughout the development of this paper.

[1] For example, Sergei Dobrovolsky, p. 309; Reuben Kessel, p. 45; Henry Latané and Donald Tuttle, p. 472; Harry Sauvain, pp. 169–73; James Van Horne, p. 76; G. Walter Woodworth, p. 191.

[2] This relationship is implicit in Macaulay's analysis of short- and long-term interest rates but is not rigorously developed. The theorem also lurks in the background in a number of other discussions of bond prices, but, to our best knowledge, has never been stated explicitly. For example, the three elements comprising the definition of duration are all identified by Homer and Leibowitz as affecting bond price volatility but are not put together. Macaulay, pp. 50–51, 61–62; Homer and Leibowitz, pp. 50–53.

is demonstrated below. It will also be demonstrated that this proposition underlies Malkiel's third theorem and that this theorem, like his second theorem, represents a special case that applies only to coupon bonds that sell at or above their par value, and to zero coupon bonds.

Duration is a concept first introduced by Frederick Macaulay to provide more complete summary information about the time structure of a bond than term to maturity.[3] Maturity provides information only about the date of final payment. However, nonzero coupon bonds generate regularly scheduled payments before maturity. Thus, maturity provides an incomplete description of the time pattern of all the payments of a bond. The longer the term to maturity, the higher the coupon rate, or the higher the market yield, the more important are the coupon payments relative to the maturity payment. Duration views a conventional nonzero coupon bond as a zero coupon serial bond with consecutive maturity payments equal to the coupon payments plus a larger payment at final maturity. It thus considers all payments generated by a bond. Duration is defined as:

$$(1) \quad D = \frac{\displaystyle\sum_{t=1}^{n} \frac{(C_t)t}{(1+r_t)^t} + \frac{An}{(1+r_n)^n}}{\displaystyle\sum_{t=1}^{n} \frac{C_t}{(1+r_t)^t} + \frac{A}{(1+r_n)^n}}$$

where

D = Duration
C = Dollar value of coupon payment
A = Dollar value of maturity payment
t = Period in which payment is made
r = Interest rate applicable for period t
n = Maturity period

This formulation is seen to be an ordinary weighted average of the time periods in which payments are to be made. Each period is weighted by the present value of the corresponding payment or price of the maturing serial bond. That is, duration identifies the length of time from the present at which the bond generates the average present value dollar. This period may be considered the average life of the bond. It is equal to the maturity of a single payment zero coupon bond selling at the same market price as the coupon bond, generating the same yield, and having a par value equal to the sum of the total payments generated by the bond when all coupons are reinvested to maturity.

For zero coupon bonds, duration is equal to maturity. For all other bonds, duration is shorter than maturity. However, the relationship between duration and maturity is non-linear and complex. For bonds priced at or above par, duration increases monotonically with maturity but at a decreasing rate. This explains both the much publicized direct relation between price volatility and term to maturity cited earlier and Malkiel's Theorem 3. For discount bonds, the relationship is more complex. Duration increases with maturity to a point before perpetuity, peaks at that point, and subsequently declines.[4] For all bonds, differences between duration and maturity are small for short maturities but increase as maturity increases. Lawrence Fisher and Roman Weil note that for all bonds duration is bounded at perpetuity by $(r+p)/rp$, where r is the yield to maturity and p is the number of times per year inter-

[3] Macaulay, pp. 45–53. A brief history of duration is traced by Fisher and Weil, pp. 415–16. Although not cited by these authors, J. R. Hicks independently of Macaulay appears to have developed the concept of duration, which he refers to as "average period" (p. 186).

[4] For discount bonds, duration achieves its maximum when

$$n = \frac{1}{r} + \frac{1}{r-c} + \frac{r}{r-c} + \frac{r-c}{cr(1+r)^n}$$

where r = interest yield to maturity, c = coupon rate, n = term to maturity, and declines when term to maturity is larger than the value of the right side of the expression. The first three terms of this equation are equivalent to the expression developed by Macaulay as an approximation of the term to maturity at which duration is at a maximum and no longer increases with increases in term to maturity. The complete expression was derived by G. O. Bierwag and may be obtained from the authors on request. As n approaches infinity, the final term approaches zero and Macaulay's approximation is obtained. See Macaulay, p. 50n. The possibility that the duration of discount bonds can be inversely related to maturity appears to have been overlooked by Malkiel, p. 57.

TABLE 1—DURATION OF SELECTED BONDS AT MARKET YIELD OF 6 PERCENT

Years to Maturity	Coupon Rate[a] .02	.04	.06	.08
	Years			
1	0.995	0.990	0.985	0.981
5	4.756	4.558	4.393	4.254
10	8.891	8.169	7.662	7.286
20	14.981	12.980	11.904	11.232
50	19.452	17.129	16.273	15.829
100	17.567	17.232	17.120	17.064
∞	17.167	17.167	17.167	17.167

[a] Coupon payments and compounding semi-annually.
Source: Fisher and Weil, p. 418.

est is paid and compounded (p. 418). For par and premium bonds, this boundary value represents maximum duration; for discount bonds it is less than maximum duration. Examples of the relationship between duration, maturity, market yield, and coupon rate are shown in Table 1 developed by Fisher and Weil. Long-term bonds are seen not to be as long as they may appear.

Duration varies inversely with the coupon interest rate. A 50-year 8 percent coupon bond yielding 6 percent is seen from Table 1 to have approximately the same duration as a 20-year 2 percent coupon bond yielding 6 percent. Fisher and Weil proceed to demonstrate the usefulness of duration in optimizing bond portfolio strategy.

The unique role of duration in relating changes in bond price to changes in interest yields may be derived by computing the differential of bond price with respect to interest rate.[5] The price of a bond (P) is equal to the sum of the present values of the stream of coupon payments and of the final payment at maturity.

$$(2) \qquad P = \sum_{t=1}^{n} \frac{C_t}{(1+r)^t} + \frac{A}{(1+r)^n}$$

Taking the differential:

$$(3) \quad dP = \left[\sum_{t=1}^{n} \frac{-(C_t)t}{(1+r)^{t+1}} - \frac{An}{(1+r)^{n+1}} \right] dr$$

[5] Other properties of duration are described by William Brock and Roman Weil, and Myron Grove.

Factoring $-1/(1+r)$ from both terms in the bracket:

$$(4) \quad dP = -\left[\sum_{t=1}^{n} \frac{(C_t)t}{(1+r)^t} + \frac{An}{(1+r)^n} \right] \frac{dr}{(1+r)}$$

If all coupon payments are discounted by the yield to maturity, the term in brackets is equivalent to the numerator in the definition of duration specified in equation (1). The denominator of equation (1) is equivalent to the definition of bond price specified in equation (2). Substituting from equations (2) and (4) into equation (1) yields:

$$(5) \qquad D = -\frac{dP}{P}\left[\frac{(1+r)}{dr}\right]$$

and solving for dP/P.

$$(6) \qquad \frac{dP}{P} = -D\frac{dr}{(1+r)}$$

For continuous discounting, r approaches 0 and equation (6) reduces to:[6]

$$(7) \qquad \frac{dP}{P} = -D\,dr$$

Equations (6) and (7) confirm the general theorem stated earlier. For a given change in interest rates, percentage changes in bond prices vary proportionately with duration. The maximum change in price occurs when the duration is at a maximum. It follows that the relationship between changes in bond prices and term to maturity depends on the relationship between duration and maturity. If, as maturity increases, duration also increases, the magnitude of the price change increases although at a declining rate. If, as maturity increases, duration decreases, the magnitude of price changes also declines.[7]

[6] These solutions were derived in a different context by Fisher, pp. 113–14.

[7] For example, in his appendix to ch. 3, Malkiel considers a series of hypothetical 2 percent coupon bonds with maturities ranging from one year to perpetuity. He observes that when the market yield is increased from 5.40 to 5.50 percent, the maximum percentage decline in price occurs for a bond with 50 years to maturity. From equation (1), this is also computed to be the approximate maturity of the bond possessing the maximum duration. Both duration and the magnitude

As the relationship between maturity and duration is complex, the relationship between maturity and bond price volatility is also complex.

The linear mathematical dependency of bond prices on duration and not on maturity has a number of important practical implications for the analysis of interest rate behavior. For example, it may be more useful to derive yield curves with respect to duration than to maturity.[8] In recent years, this would show longer maturity Treasury securities to have approximately similar average lives but greatly different before-or after-tax yields. Likewise, it may be necessary to reconsider the contents of the information obtained from forward rates conventionally calculated from yield curves derived from observed coupon securities. It is unlikely that a rate computed for two coupon bonds of different maturities but similar durations indicates the rate an investor would pay now for a forward security commencing at the maturity of the shorter security and maturing at the maturity of the longer security. This may help to explain the limited success of empirical tests in differentiating among competing theories of the term structure.

Lastly, the use of maturity rather than of duration to compare securities may affect the empirical measurements of both liquidity and default risk premiums. The higher interest rates are, the shorter is duration. Thus, the absolute difference in duration between a given pair of short- and long-term bonds will be greater when interest rates are low than when rates are high. As a result, the finding that in a symmetrical cycle yields on long-term securities exceed yields on short-term securities when yields are low by more than short-term yields exceed long-term yields when yields are high may in part or in total reflect the failure to maintain constant the difference in duration.[9] This measurement error may also help explain the inverse relationship between liquidity premiums and the level of interest rates reported by some investigators.[10]

The inverse relationship between duration and coupon makes a higher coupon bond a shorter term bond than a lower coupon bond of the same maturity. To the extent bonds possessing higher risks of default carry higher coupons than default free bonds, a comparison of the rates on these two bonds of equal maturity will underestimate the magnitude of the default risk premium in periods of upward sloping yield curves and overestimate the premium in periods of downward sloping yield curves.[11]

REFERENCES

W. Brock and R. L. Weil, *The Axiomatics of Macaulay's Duration*, Univ. of Chicago, May 1971.

S. P. Dobrovolsky, *The Economics of Corporation Finance*, New York 1971.

L. Fisher, "An Algorithm for Finding Exact Rates of Return," *J. Bus. Univ. Chicago*, Jan. 1966, *39*, 111–18.

———, "Determinants of Risk Premiums on Corporate Bonds," *J. Polit. Econ.*, June 1959, *67*, 217–37.

——— and R. L. Weil, "Coping With the Risk of Interest Rate Fluctuations: Returns to Bondholders From Naive and Optional Strategies," *J. Bus. Univ. Chicago*, Oct. 1971, *44*, 408–31.

W. C. Freund, *Investment Fundamentals*, Washington 1970.

M. A. Grove, "A Model of the Maturity Profile of the Balance Sheet," *Metroecon.*, Jan.-Apr. 1966, *13*, 40–55.

J. R. Hicks, *Value and Capital*, Oxford 1939.

S. Homer and M. Leibowitz, *Inside the Yield Book*, Englewood Cliffs 1971.

R. A. Kessel, *The Cyclical Behavior of the*

of the price decline diminish as the maturity of the discount bonds increase beyond this point. See Malkiel, p. 80.

[8] This is suggested by Macaulay, pp. 44, 50–51, 51–62, 68.

[9] See Kessel, pp. 81–90. To the extent liquidity premiums are a linear function of the price riskiness of bonds and price risk is a linear function of duration, liquidity premiums are also a linear function of duration. As, in turn, duration increases with term to maturity but at a decreasing rate, this explains the hypothesis that liquidity premiums increase with term to maturity at a decreasing rate. See Malkiel, pp. 24–28.

[10] See Kessel, pp. 25–28; Van Horne, pp. 347–50.

[11] See Fisher, pp. 217–37.

Term Structure of Interest Rates, New York 1965.

H. A. Latané and D. L. Tuttle, *Security Analysis and Portfolio Management,* New York 1970.

F. R. Macaulay, *Some Theoretical Problems Suggested by the Movements of Interest Rates, Bond Yields and Stock Prices in the U.S. Since 1856,* New York 1938.

B. G. Malkiel, *The Term Structure of Interest Rates,* Princeton 1966.

H. Sauvain, *Investment Management,* Englewood Cliffs 1967.

J. C. Van Horne, *Function and Analysis of Capital Market Rates,* Englewood Cliffs 1970.

———, "Interest-Rate Risk and the Term Structure of Interest Rates," *J. Polit. Econ.,* Aug. 1965, *73,* 344–51.

G. W. Woodworth, *The Money Market and Monetary Management,* 2d ed., New York 1972.

The End of the North-South Wage Differential: Comment

By MARK L. LADENSON*

In a recent issue of this *Review*, Philip Coelho and Moheb Ghali have argued that various studies purporting to show a persistent differential between wages in northern and southern states have used data unadjusted for regional differences in price level; that the differential has been proved for nominal wages, not real wages. Using data for 1963 they show that the average nominal wage rate for a sample of five northern metropolitan areas is significantly greater than such an average for five "southern" metropolitan areas. However, after the nominal wage figures are adjusted for differences in the cost of living in each of the ten areas, this differential disappears. This result survives the introduction of dummy variables to account for differences in industry composition, sex, color, and capital-labor ratio.

It seems almost unbelievable that earlier studies of the wage differential have ignored regional differences in price levels, and Coelho and Ghali must be commended for calling attention to this anomaly. However, they have not proved the end of the North-South wage differential; at least not to the satisfaction of this writer. The reason is that of the five cities used in their southern sample, only one, Atlanta, would be accepted by many as southern, the classification of Baltimore, Dallas, Houston, and Washington, D.C. as southern by the Departments of Labor and Commerce notwithstanding.[1] The purpose of this note is to repeat Coelho and Ghali's comparisons for another sample of cities, half of whose members would be recognized as southern by any reasonable observer.[2] Required data for each metropolitan area are number of production workers, total man-hours, and total wages for production workers, classified by industry, and cost of living indices. The first three items are provided in the *Census of Manufactures 1967* and the latter can be obtained for the year 1967 from the *Handbook of Labor Statistics 1970*.

Table 1 shows that our sample seems to give roughly the same results as that of Coelho and Ghali. Average hourly nominal wages for the five northern areas are 14.4 percent higher than for the southern areas. This differential drops to 4 percent when real hourly wages are considered. The same phenomenon occurs for annual wages. Looking at the figures for the individual cities however, it is clear that Baton Rouge is an extreme outlier. If we compare average hourly wages between the five northern areas and the four southern areas excluding Baton Rouge, we find that when we shift from nominal to real wages the percentage difference drops from 25.6 to 15.1, still a sizeable difference.

It would be preferable to establish the continued existence of the North-South wage differential without excluding Baton Rouge from the sample. This might be accomplished by showing that the high average wages in Baton Rouge are a result of the industrial composition of that region. Our method of distinguishing between regional effects and industry effects is identical to that of Coelho and Ghali. We calculate wage rates and annual wages for every industry (two digit classification) and for every metropolitan area.[3] The model takes the form:

* Assistant professor of economics, Michigan State University.

[1] The title of one of the sessions at the December 1971 meeting of the American Economic Association, at which two papers were presented, was "Is the South Still Backward?" M. I. Foster included only Atlanta in his definition of the South. F. Ray Marshall omitted Baltimore and Washington, D.C. from his definition.

[2] We do not repeat the last part of their analysis involving regional differences in sex, color, and capital-labor ratio.

[3] No observations were available for the following industries in: Boston, SIC 21; Buffalo, SIC 21, 31; Pitts-

TABLE 1—MONEY WAGES, ANNUAL WAGES, REAL WAGES, AND REAL ANNUAL WAGES FOR SELECTED METROPOLITAN AREAS, 1967

Metropolitan Area	Hourly Money Wage	Annual Money Wage	Cost of Living Deflator	Hourly Real Wage	Real Annual Wage
Boston	\$2.94	\$5734.9	106	\$2.77	\$5410.3
Buffalo	3.47	7013.2	103	3.37	6808.9
New York & NE New Jersey	2.86	5569.2	102	2.80	5460.0
Philadelphia	3.07	6051.8	100	3.07	6051.8
Pittsburgh	3.55	6850.3	99	3.59	6919.4
Simple Average (North)	\$3.18	\$6243.9		\$3.12	\$6130.1
Atlanta	\$2.80	\$5670.1	95	\$2.95	\$5968.5
Baton Rouge, La.	3.75	7700.9	91	4.12	8462.5
Durham, N. C.	2.43	4740.5	94	2.58	5043.1
Nashville	2.49	5028.5	93	2.68	5407.0
Orlando, Fla.	2.45	5067.4	92	2.66	5580.4
Simple Average (South)	\$2.78	\$5641.5		\$3.00	\$6092.4
Simple Average (South; Baton Rouge omitted)	\$2.53	\$5126.6		\$2.71	\$5499.8

$$W_{ij} = \alpha + \beta_i + \gamma_j$$

where W_{ij} is the wage in region i and industry j, α is the intercept, β_i is the region effect, and γ_j is the industry effect. The wage W_{ij} is regressed on twenty dummies, the first of which assumes a value of unity for northern metropolitan areas and zero for southern metropolitan areas. The other nineteen dummies each assume a value of unity for one of the twenty industries and zero for every other industry with industry SIC 20 as a base.

It is to be emphasized that observations on Baton Rouge *were* included in the regressions. Looking at Table 2 we see why Baton Rouge is such an extreme outlier. In addition to SIC 20, observations are available for Baton Rouge on SIC 27, 28, 32, 34, implying that these are the major industries in the area. In all four columns of Table 2 the coefficients of these industries are significant and large.

burgh, SIC 21, 22, 24, 31; Atlanta SIC 21, 29, 38; Nashville, SIC 21, 29, 31, 33, 38; Orlando, SIC 21, 22, 26, 29, 30, 31, 33, 37, 38, 39. Observations were available only for the following industries in Baton Rouge, SIC 20, 27, 28, 32, 34; Durham, SIC 20, 22, 27, 32, 36. Instead of 200 observations (20 industries in each of 10 cities) we therefore have 145.

Turning to the coefficient of the North-South dummy we find these also to be large and significant in the first two columns. The fifty-five cent differential between northern and southern nominal hourly wages implies that these are 24.2 percent lower in the South than over the entire sample. The \$991.70 differential between northern and southern nominal annual wages implies that these are 21.2 percent lower in the South than in the sample taken as a whole. These differences are reduced but by no means vanish when we consider real wages. The thirty-two cent North-South difference in hourly real wages means these are 13.1 percent higher in the North than in the entire sample. The \$529 difference in annual real wages means these are 10.6 percent higher in the North than in the entire sample.

Our conclusion is as stated in the second paragraph. Coelho and Ghali have done a service by calling attention to the necessity of using data on *real* wages when considering regional wage differentials. However, the use of such data indicate that the North-South wage differential still exists (or existed in 1967) when the cities used for southern observations in the sample belong unambiguously in that category.

Table 2—Money Wages and Incomes, Real Wages and Incomes: Regression Results[a]

Variable	Annual Money Wage Coef.	Annual Money Wage *t*	Hourly Money Wage Coef.	Hourly Money Wage *t*	Annual Real Wage Coef.	Annual Real Wage *t*	Hourly Real Wage Coef.	Hourly Real Wage *t*
Intercept	4661.5	*23.5*	2.271	*22.39*	5009.8	*23.3*	2.440	*21.96*
N-S dummy	991.7	*9.1*	.550	*9.85*	529.0	*4.5*	.320	*5.22*
Dummy of Industry								
SIC 21	−1307.0	*−2.8*	−.611	*−2.53*	−1241.7	*−2.4*	−.574	*−2.18*
22	−611.5	*−2.1*	−.324	*−2.13*	−658.5	*−2.0*	−.344	*−2.07*
23	−1619.6	*−5.7*	−.606	*−4.14*	−1549.7	*−5.3*	−.615	*−3.83*
24	−763.6	*−2.6*	−.362	*−2.38*	−797.6	*−2.5*	−.380	*−2.28*
25	−332.9	*−1.2*	−.162	*−1.11*	−340.2	*−1.1*	−.167	*−1.04*
26	62.8	*0.2*	.013	*0.08*	46.6	*0.1*	.011	*0.07*
27	1058.2	*3.9*	.738	*5.35*	1098.2	*3.8*	.766	*5.07*
28	1311.6	*4.7*	.626	*4.42*	1384.2	*4.6*	.661	*4.26*
29	2224.2	*6.6*	.862	*5.03*	2186.6	*6.0*	.858	*4.58*
30	59.9	*0.2*	.013	*0.08*	49.6	*0.2*	.009	*0.05*
31	−1365.4	*−3.8*	−.542	*−2.96*	−1403.0	*−3.6*	−.560	*−2.79*
32	467.2	*1.7*	.301	*2.18*	475.2	*1.6*	.313	*2.07*
33	1430.6	*4.6*	.738	*4.60*	1422.4	*4.2*	.738	*4.20*
34	736.7	*2.7*	.414	*2.92*	769.2	*2.6*	.435	*2.80*
35	1074.4	*3.8*	.474	*3.24*	1077.4	*3.5*	.480	*2.99*
36	381.1	*1.4*	.176	*1.24*	372.7	*1.2*	.172	*1.10*
37	1874.0	*6.3*	.903	*5.92*	1872.5	*5.8*	.904	*5.41*
38	446.9	*1.3*	.233	*1.36*	447.8	*1.2*	.236	*1.26*
39	−888.1	*−3.0*	−.297	*−1.95*	−896.1	*−2.8*	−.297	*−1.78*
$\bar{R}^2$	.764		.736		.705		.661	

[a] The industries corresponding to the Standard Industrial Classification Code are:

SIC 20 = Food and Kindred Products
21 = Tobacco
22 = Textile Mill Products
23 = Apparels
24 = Lumber and Wood
25 = Furniture and Fixtures
26 = Paper and Allied Products
27 = Printing and Publishing
28 = Chemicals
29 = Petroleum and Coal
30 = Rubber and Plastics
SIC 31 = Leather
32 = Stone and Clay
33 = Primary Metal
34 = Fabricated Metal
35 = Nonelectrical Machinery
36 = Electrical Machinery
37 = Transportation Equipments
38 = Instruments
39 = Miscellaneous

REFERENCES

P. R. P. Coelho and M. A. Ghali, "The End of the North-South Wage Differential," *Amer. Econ. Rev.*, Dec. 1971, *61*, 932–37.

M. I. Foster, "Is the South Still a Backward Region and Why?" *Amer. Econ. Rev. Proc.*, May 1972, *62*, 195–203.

F. R. Marshall, "Some Rural Economic Development Problems in the South," *Amer. Econ. Rev. Proc.*, May 1972, *62*, 204–11.

U.S. Bureau of the Census, *Census of Manufactures: 1967*, Washington 1970.

U.S. Department of Labor, *Handbook of Labor Statistics 1970*, Bull. No. 1666, Washington 1970.

The End of the North-South Wage Differential: Reply

By PHILIP R. P. COELHO AND MOHEB A. GHALI*

In his comment, Mark Ladenson contends that substantial differences in wages still exist between the North and the South even after accounting for differences in regional prices. We find his results objectionable on three grounds: a) his arbitrary exclusion of available data, b) his choice of a low budget cost of living as a deflator, and c) the possibility of heteroscedasticity which would render his testing procedure invalid.

Ladenson objects to four of the five cities we used as our southern sample, all previous empirical work notwithstanding. These five cities were used in our analysis because they were the only five cities for which the necessary 1963 data were available. Similarly, the five cities included in our Northeast sample were the only Standard Metropolitan Statistical Areas (*SMSA*) for which complete 1963 data existed.[1] If we consider 1967 data, there are many more *SMSA*s in both regions than are utilized by Ladenson. (Our data sources are identical to Ladenson's: *Census of Manufacturers 1967* and *Handbook of Labor Statistics 1970*.) In this study we use all the data available by: 1) examining wage differentials throughout the entire United States, 2) offering alternative definitions for the Northeast and South, and 3) using different deflators to measure differentials in the regional cost-of-living.

* Assistant professor of economics, Washington State University, and associate professor of economics, University of Hawaii, respectively.

[1] In our original paper (hereafter C-G) the geographical term North was defined to include the New England and Middle Atlantic states, the region defined by the Bureau of the Census as the Northeast. Because we are convinced that a finer use of geographical terms is necessary to avert terminological debates and to avoid confusion in the examination of wage differentials in other parts of the country, we have modified our regional definition. In this paper unless explicitly stated otherwise, we accept the Bureau of the Census' definitions for the four major regions of the country (Northeast, North Central, South, and West).

The statistical techniques used are similar to those employed by Ladenson and in our previous paper, p. 934, except that we have added two additional dummy variables to account for the additional geographical regions, the North Central and western regions of the United States. Dummy variables, then, exist for the South, North Central, and West of the United States; consequently we are comparing wages in these regions to wages in the Northeast.[2] Industry 21 of the Standard Industrial Classification

[2] The cities for which two-digit industrial classification and comparative cost-of-living data exist are the following:

Northeast Region:
Boston, Mass., Buffalo, N. Y., Hartford, Conn., Lancaster, Pa., New York City and northeastern New Jersey, Philadelphia, Pa., Pittsburgh, Pa., Portland, Me.,

Southern Region:
Atlanta, Ga., Austin, Texas, Baltimore, Md., Baton Rouge, La., Dallas, Texas, Durham, N. C., Houston, Texas, Nashville, Tenn., Orlando, Fla., Washington, D.C.

North Central Region:
Cedar Rapids, Ia., Champaign-Urbana, Ill., Chicago-Northwestern Indiana, Cincinnati (Ohio, Kentucky, Indiana), Cleveland, Ohio, Dayton, Ohio, Detroit, Mich., Green Bay, Wis., Indianapolis, Ind., Kansas City, Mo., Milwaukee, Wis., Minneapolis-St. Paul, Minn., St. Louis, Mo., Wichita, Kan.,

Western Region:
Bakersfield, Ca., Denver, Col., Honolulu, Hawaii, Los Angeles-Long Beach, Ca., San Diego, Ca., San Francisco-Oakland, Ca., Seattle-Everett, Wash.

Ladenson stresses that Baton Rouge, a high wage city, is included in his regression results for the South. Baton Rouge is a high wage city because of its industrial mix, as he points out. Consequently, including Baton Rouge, and using the econometric techniques which both Ladenson and we use, can have no effect upon the southern coefficient unless wages in the Baton Rouge industrial mix are different from wages in the Northeast for the same industrial mix. Ladenson's point is specious.

TABLE 1—ORDINARY LEAST SQUARES
(Coefficients and t-values)

	Sample I		Sample II		Sample III	
	e	*w*	*e*	*w*	*e*	*w*
Money						
R^2	.66	.65	.67	.66	.68	.67
Int	5817	2.87	5862	2.89	5767	2.83
NC	173.02	.09*	172.69*	.09*	292.75*	.16*
	1.96	*2.17*	*1.99*	*2.23*	*3.03*	*3.51*
S	−621.76*	−.34*	−828.52*	−.46*	−711.16*	−.39*
	−6.32	*−7.25*	*−6.69*	*−7.83*	*−5.40*	*−6.28*
W	235.07*	.18*	230.10*	.18*	347.68*	.25*
	2.23	*3.69*	*2.22*	*3.75*	*3.09*	*4.74*
Low Budget						
R^2	.63	.62	.64	.63	.64	.63
Int	5687	2.80	5727	2.82	5609	2.75
NC	261.56*	.13*	261.14*	.13*	403.69*	.21*
	2.90	*3.19*	*2.92*	*3.22*	*4.04*	*4.64*
S	−241.51*	−.15*	−354.33*	−.22*	−215.38	−.14*
	−2.41	*−3.20*	*−2.77*	*−3.73*	*−1.59*	*−2.27*
W	12.51	.06	7.55	.05	147.66	.14*
	.12	*1.29*	*.07*	*1.27*	*1.27*	*2.66*
Intermediate Budget						
R^2	.63	.62	.64	.63	.65	.64
Int	5586	2.75	5630	2.78	5487	2.69
NC	371.78*	.19*	371.44*	.19*	538.06*	.28*
	4.07	*4.42*	*4.11*	*4.48*	*5.36*	*6.06*
S	−25.13	−.04	−175.29	−.13*	−12.51	−.04
	−.25	*−.89*	*−1.36*	*−2.21*	*−.09*	*−.66*
W	241.91*	.18*	236.86*	.18*	401.28*	.27*
	2.22	*3.55*	*2.20*	*3.57*	*3.44*	*5.04*
High Budget						
R^2	.63	.62	.64	.63	.65	.64
Int	5591	2.75	5633	2.78	5470	2.68
NC	454.95*	.23*	454.64*	.23*	643.66*	.33*
	4.94	*5.37*	*4.99*	*5.43*	*6.38*	*7.19*
S	−60.47	−.06	−189.40	−.14*	−4.59	−.04
	−.59	*−1.24*	*−1.45*	*−2.29*	*−.03*	*−.59*
W	215.32*	.17*	210.48	.17*	397.46*	.27*
	1.96	*3.27*	*1.94*	*3.28*	*3.39*	*4.99*

Note: *Int* = Intercept
NC = North Central Region
S = Southern Region
W = Western Region
t-values are in italics below the coefficients.
Starred entries indicate that the coefficients are significantly different from zero at the .05 level.

Code (SIC) is dropped for lack of sufficient data; industry 20 of the SIC serves as a base.

Our results are reported in Table 1. Only regional coefficients are reported.[3] Sample I includes all the *SMSA*s for which 1967 data are available. Sample II excludes the four southern cities to which Ladenson objects and Austin, Texas. Sample III has the same coverage as Sample II except that two northeastern *SMSA*s, Pittsburgh and Buffalo, are excluded. For each sample we report the results for hourly wages (w) and annual earnings (e). The results for money wages

[3] The results of the entire set of regressions are available from the authors upon request.

and earnings are reported in the first five rows. The remaining rows report the results for real wage and real earnings when three alternative deflators are used: Low, Intermediate, and High Budget cost-of-living indexes.

For money wages Ladenson finds that, after accounting for industry mix, annual wages in his "South" are 17.5 percent below the Northeast.[4] In a recent study of interstate wage differentials Gerald Scully found that money wages in the South (Census Bureau definition) were 11.6 percent below those in the Northeast, p. 768. In our original study, p. 934, we found annual wages in money terms in the South to be 10.5 percent below those in the North, hourly wages were 12.4 percent below those in the North. In the present study, in Sample I, the South is 10.7 percent and 11.7 percent below the North for annual money wages and hourly earnings respectively. Whose definition of the South should be accepted?

First, it appears that there should be no dispute over our inclusion in the Northeast of Hartford, Connecticut, Portland, Maine, and Lancaster, Pennsylvania. Complete data for these cities are available for 1967, Ladenson did not use them. Secondly, if Ladenson's results are accepted, one must conclude that between 1958 and 1967 the position of southern industrial workers deteriorated relative to their northeastern counterparts. All the evidence available, quantitative and subjective, belies this conclusion.

In Sample I, for real wages and real earnings the coefficients of the South are insignificant at the 5 percent level except when the data are deflated by the low budget index. We are faced with another problem, which index should we use? Quite obviously, the high budget index is inappropriate for most industrial workers' families.[5] The choice between the low and intermediate indexes is not so clear cut. The annual dollar earnings of each worker would maintain a low budget standard of living. However, there is more than one worker per family; in 1967 50.3 percent of all families in the United States had two family members in the labor force. Furthermore, our data are obtained by dividing the total annual wages paid by the total labor force. No allowance is made for overtime, or part-time and seasonal workers. We believe that the intermediate budget deflator is subject to less criticisms than the alternatives, consequently our conclusions about the difference in wages between the Northeast and South are based upon the use of this index.[6]

In Sample II we drop the cities in areas which Ladenson believes are nonsouthern. However, we use all the northeastern data which are available. The differences between our results and Ladenson's are instructive; in annual money wages and hourly earnings Ladenson has the South 17.5 percent and 19.5 percent below the Northeast, respectively. Our results for the identical variables have the South 14.1 percent and 15.7 percent below the Northeast. Simply by excluding three cities from the Northeast, the differential is widened. Here lies the danger of arbitrarily discarding data.

Ladenson does not give any rationale for rejecting the Census Bureau's definition of the South. By redefining the South, Ladenson's work is not comparable to other studies of the South; consequently, his conclusions are suspect because his data base is different and severely limited. Ladenson's definition of the South has imposed severe limitations upon his sample, resulting in both regional and industrial composition imbalances. Out of his sample of 145 observations only about one-third represent observations on southern industries (52 observations). The imbalance in the distribution of industries may be the

[4] Ladenson states that southern wages "... are 24.2 percent lower in the South than over the entire sample." He actually means that with his data the North is 24.2 percent higher than the South. The South is his base, we have changed his results to a northeastern base to make them comparable with our present and previous findings.

[5] The average annual dollar cost for maintaining a higher-income standard of living for the United States in 1967 was $13,367. This statistic (and others in this section) were obtained from the *Handbook of Labor Statistics, 1970.*

[6] Ladenson uses a low budget deflator although he never mentions it. This is another area of disagreement between us.

TABLE 2—GENERALIZED LEAST SQUARES
(Coefficients and *t*-values)

	Sample I		Sample II		Sample III	
	e	*w*	*e*	*w*	*e*	*w*
Money						
R^2	.87	.87	.87	.87	.87	.87
NC	1258.89*	.60*	1183.43*	.56*	1363.54*	.66*
	7.80	*7.36*	*7.03*	*6.63*	*7.53*	*7.32*
S	384.79	.11	−205.76	−.19	28.86	−.06
	1.37	*.75*	*−.41*	*−.75*	*.06*	*−.25*
W	1224.93*	.60*	1139.68*	.56*	1300.60*	.65*
	5.59	*5.49*	*5.00*	*4.92*	*5.48*	*5.50*
Low Budget						
R^2	.87	.86	.87	.87	.87	.87
NC	1274.88*	.60*	1201.51*	.57*	1392.35*	.68*
	7.98	*7.54*	*7.21*	*6.80*	*7.78*	*7.58*
S	713.19*	.27	213.58	.02	453.03	.15
	2.57	*1.91*	*.43*	*.08*	*.90*	*.58*
W	861.39*	.42*	777.96*	.38*	946.17*	.47*
	3.98	*3.86*	*3.45*	*3.36*	*4.04*	*4.04*
Intermediate Budget						
R^2	.87	.87	.87	.87	.87	.87
NC	1443.17*	.69*	1371.48*	.65*	1602.87*	.78*
	9.31	*8.86*	*8.49*	*8.07*	*9.32*	*9.14*
S	1047.84*	.43*	530.86	.13	812.83	.32
	3.88	*3.19*	*1.11*	*.73*	*1.68*	*1.34*
W	1242.08*	.61*	1159.75*	.57*	1362.66*	.68*
	5.91	*5.80*	*5.30*	*5.22*	*6.05*	*6.09*
High Budget						
R^2	.87	.87	.87	.87	.88	.88
NC	1588.53*	.76*	1517.80*	.73*	1778.75*	.87*
	10.40	*9.94*	*9.53*	*9.11*	*10.57*	*10.39*
S	1045.44*	.43*	586.08	.20	902.88	.37
	3.93	*3.23*	*1.24*	*.85*	*1.90*	*1.56*
W	1230.14*	.61*	1149.49*	.57*	1378.99*	.69*
	5.94	*5.84*	*5.33*	*5.25*	*6.25*	*6.30*

Note: *NC* = North Central *W* = West
S = South *t*-values are in italics below the coefficients.
Starred entries are significant at the .05 level.

cause of his unusually high differentials in nominal wages. If one considers the sign of the industry effects reported in Ladenson's Table 2, seven missing observations for the northern sample are all low wage industries; on the other hand, 30 out of the 48 missing observations for the southern sample have positive signs.

One of the statistical pitfalls in Ladenson's analysis is the inclusion of industries in his northern sample which have no counterpart in his South. The value of the model used is to separate industry effects from regional effects, and, consequently, adjust the regional wage for industry mix. This objective is not achievable if the industry exists in one region only. Industries which are missing from Ladenson's South are SIC 21, 29, and 38. Two other observations (SIC 31 and 33) have only one observation apiece in the southern sample. For this reason we dropped SIC 21 from the sample in our previous

paper, p. 934, and for the same reason we dropped it from the present analysis. With respect to the two other industries, SIC 29 and 38, which would not be represented in the southern sample if we follow Ladenson's definition as we do in Samples II and III, the decomposition of wages to industry and regional effects is accomplished by including the other two regions of the country, where these industries exist.

Nevertheless, let us accept, for illustration, Ladenson's definition and, again for illustration, define the Northeast as the area north of the Mason-Dixon line and east of the Appalachian Mountains. This definition could be defended in terms of geography, history, and market orientation; it constitutes our Sample III. Except for hourly earnings deflated by the low standard of living index, all the southern coefficients are insignificantly different from zero at the 5 percent level. We do not advocate regional definitions of the South nor of the Northeast as defined in Sample III, we do wish to stress that there can be serious biases when data are arbitrarily discarded from a sample.

The results reported may be biased, and, therefore, misleading on two further counts. First, the results may be subject to a specification bias which results from the omission of pertinent variables (such as sex, color, capital-labor ratio). The omitted variables may be working in the direction of concealing the true wage differential. This possibility was investigated in our earlier paper, p. 936, and the conclusion was that inclusion of the omitted variables did not affect our results. This procedure cannot be followed in the present study because the 1958 data on capital and the 1960 data on sex and color are inappropriate for analyzing 1967 wage data.

The second possibility of bias would result if our estimated variance-covariance matrix is biased. Suppose the wage equation which holds for the individual worker k is $w_{ij}^k=\alpha+\beta_i+\gamma_j+e_{ij}^k$ where i represents the ith industry, j the jth city, α the intercept, β_i the industry effects, and γ_j regional effects. If we make the assumption that the error terms e_{ij}^k are spherical normal, then the error term in the equation pertaining to average industry wage levels will not be homoscedastic:

$$w_{ij} = \frac{1}{n_{ij}} \sum_k w_{ij}^k = \alpha + \beta_i + \gamma_j + \frac{1}{n_{ij}} \sum_k e_{ij}^k$$

where n_{ij} represents the number of workers in the ith industry in the jth city. Under these conditions, estimated standard errors of the regression coefficients will be biased, and the calculated t-statistics will not be distributed t, when the results are obtained by applying ordinary least squares. The application of generalized least squares, where the variance matrix is a diagonal matrix with elements $1/n_{ij}$, consists of weighting the original data by a factor which will correct for heteroscedasticity. In our case, the correction amounts to weighting our data by $\sqrt{n_{ij}}$.

In Table 2 the generalized least square results are shown for each of the three samples. Table 2 shows that, in money terms, wages in the South (both annual and hourly) are insignificantly different from wages in the Northeast.[7] Using generalized least squares and deflating by a cost of living index reveals even more startling results. All the southern coefficients are positive and some of them are significant. It is of interest to point out that the southern coefficients are the only ones which change their signs. The other regional coefficients do become larger and their t-values increase. After we account for heteroscedasticity, the Northeast is the low real-wage region.

In conclusion, the results of our previous paper are confirmed and there exist no significant differences in real wages between the Northeast and the South when the accepted

[7] A straightforward interpretation in dollar magnitudes of the regression coefficients is not possible using the correction described. Thus, we have not reported intercept values, and we caution the reader not to interpret the values in simple dollar terms as can be done in Table 1. Complete regression results are available from the authors upon request.

definition of the South is used and when money wages are deflated by the intermediate cost of living index. There are wage differentials within the United States, however. The North Central and western regions have significantly higher wages than the Northeast. Whether these differentials can be accounted for by some other explanatory variables, or whether they are equalizing differentials for nonpecuniary regional differences or whether the economy was simply in disequilibrium in 1967 are questions which can only be answered by further research.

REFERENCES

P. R. P. Coelho and M. A. Ghali, "The End of the North-South Wage Differential," *Amer. Econ. Rev.*, Dec. 1971, *61*, 932–37.

G. W. Scully, "Interstate Wage Differentials: A Cross Section Analysis," *Amer. Econ. Rev.*, Dec. 1969, *59*, 757–73.

U.S. Bureau of the Census, *Census of Manufacturers: 1967*, Washington 1970.

U.S. Department of Labor, *Handbook of Labor Statistics 1970*, Bull. No. 1666, Washington 1970.

Estimation of Structure-Profit Relationships: Comment

By JOHN M. VERNON AND MARJORIE B. MCELROY*

Recently, the Federal Trade Commission (*FTC*) has accused breakfast cereal manufacturers of illegal monopolization (1971). A major basis for the charge is the *FTC* belief that high advertising expenditures by the companies create effective entry barriers. One *FTC* study (1969) reached this conclusion after finding that the coefficient of the advertising/sales ratio in a multiple regression equation explaining firm profit rates was positive and significant. Several other of these so-called market structure-performance studies have recently been completed (Marshall Hall and Leonard Weiss, Vernon (1972)). However, in a recent issue of this *Review*, Blake Imel and Peter Helmberger (hereafter, I-H) presented some findings which cast serious doubt upon the validity of these studies. According to their carefully specified profit equation for a diversified firm, the significance of advertising as a barrier to entry depends critically on the size of an unknown parameter termed the "*omega* ratio." For example, for one specification of the profit rate equation, the estimated t-ratios varied from 3.31 to 1.86 to 1.29 as the *omega* ratio ranged from 0.0 to 0.5 to 0.8. Since I-H were unable to estimate this ratio and it is not known a priori, it is important to seek evidence as to its magnitude.

In this paper, after a brief review of the I-H model, we suggest a method for estimating the *omega* ratio. We then present some estimates for a body of data similar to that used by I-H. Our results indicate that the estimated ratio (and hence the significance of the estimated coefficients) is approximately the same for two alternative specifications of the model. More importantly, the estimated advertising coefficients, conditional on our estimated *omega* ratios, are positive and statistically significant. Thus, this note tends to corroborate the hypothesis that advertising serves as an effective barrier to entry.

I. The Imel-Helmberger Model

At the cost of placing severe restrictions on the profit equation of a diversified firm, I-H are able to specify the off-diagonal as well as the diagonal elements of the error variance-covariance matrix. Hence, except for the problem of estimating the *omega* ratio and therefore the variance-covariance matrix, generalized least squares (*GLS*) is the appropriate estimation technique.

For a firm specialized in one market, I-H write the profit rate equation as

$$R_i = a + bE_i + cM_i + u_i \tag{1}$$

where

R_i = the profit rate of the ith firm
M_i = a market-related structural variable, such as the concentration ratio, of the ith firm
E_i = a firm-related variable, such as size, of the ith firm
u_i = the total error term, representing both firm-related and market-related omitted variables

A distinctive feature of their model is the assumption that the error term is composed of two components,

$$u_i = e_i + m_i$$

where e_i is that part of the error term due to the omission of firm-related variables and m_i is that part due to the omission of market-related variables. It is assumed that

$$Eu_i = Em_i = Ee_i = 0$$

and

$$\sigma_u^2 = \sigma_m^2 + \sigma_e^2$$

* Professor and assistant professor of economics, Duke University, respectively. We wish to thank Michael McElroy for valuable criticism, Terry Seaks for research assistance, and Bob Nourse for help in collecting the data. We are also grateful to the Marketing Science Institute for its support of the data collection effort.

I-H construct the profit equation for the ith diversified firm from the simple sum of profits of its divisions. Thus the profit equation of the ith diversified firm consisting of two divisions—the gth division operating in market j and the kth division operating in market h—is given by

$$(2)\quad R_{id} = a + b(D_{ij}M_j + D_{ih}M_h) + c(D_{ij}E_{gj} + D_{ih}E_{kh}) + D_{ij}m_j + D_{ih}m_h + D_{ij}e_{gj} + D_{ih}e_{kh}$$

where

D_{ij} = the share of the diversified firm's total sales accounted for by its sales in the jth market

D_{ih} = the share of the diversified firm's total sales accounted for by its sales in the hth market

Generalized to T markets the structure of the variance-covariance matrix of the error term resulting from this model is of special importance. The ith diagonal element is:

$$(D_{i1}^2 + D_{i2}^2 + \ldots + D_{iT}^2)\sigma_u^2$$

The expression in parenthesis is a Herfindahl measure of product diversification, or H_i. For a firm specialized in one market, $H_i = 1$ and hence σ_u^2 is the variance of the error term. A typical off-diagonal element is simply the covariance term for two firms. For example, if firms i and j have at least one market in common, the element is the sum of the cross-products of D_{ik} for the firms' common markets.

$$\sigma_m^2 \sum_{k=1}^{T} D_{ik}D_{jk}$$

If σ_u^2 is factored out of the variance-covariance matrix, the diagonal elements become the H measures of product diversification for the firms in the sample. The typical off-diagonal element becomes

$$K \sum_{k=1}^{T} D_{ik}D_{jk}$$

where K is the *omega* ratio, $K = \sigma_m^2/\sigma_u^2$. Hence, knowledge of K and D_{ik} is sufficient to compute the entire variance-covariance matrix.

II. Estimation Procedure

Our procedure for estimating the *omega* ratio is a simple adaptation of the two-stage technique described by Marc Nerlove, p. 373. In the first stage we account for the effect of the omitted market variables by treating m_k as parameters, and by estimating them using weighted least squares (WLS). These in turn are used to estimate σ_m^2 and thereby K and the variance-covariance matrix. Our procedure differs from that of Nerlove in that the I-H specification requires weighted dummies as well as weighted least squares. The second stage consists of using the estimated variance-covariance matrix and applying generalized least squares. The remainder of this section will be devoted to a fuller discussion of the first stage.

Generalizing the profit rate equation for the ith diversified firm selling in up to T markets yields:

$$(3)\quad R_{id} = bM_i + cE_i + m_1D_{i1} + \ldots + m_TD_{iT} + v_i$$

where

$$M_i = \sum_{k=1}^{T} D_{ik}M_k, \qquad E_i = \sum_{k=1}^{T} D_{ik}E_{ik},$$

$$v_i = \sum_{k=1}^{T} D_{ik}e_{ik}$$

with

$$Ev_i = 0$$

$$\text{var } v_i = \sigma_e^2 \sum_k D_{ik} = \sigma_e^2 H_i$$

Regarding m_k as parameters to be estimated, equation (3) can be estimated using weighted least squares[1] where the appropriate weights are $1/\sqrt{H_i}$.

[1] In Nerlove's first stage, the role of D_{ik} in equation (3) is played by a zero-one dummy variable. In the context of the I-H model, such zero-one dummies would be appropriate only if firms were specialized in one market in which case the coefficient of a dummy variable (plus the constant) would be the profit rate intercept for that market. However, the profit rate intercept for a diversi-

Applying *WLS* to (3) yields an estimated intercept term $\hat{m}_k$ for each of T markets. Thus our estimator of the variance of the error terms due to omitted market variables is given by

$$\hat{\sigma}_m^2 = \frac{1}{T}\sum_{k=1}^{T}\left(\hat{m}_k - \frac{1}{T}\sum_{k=1}^{T}\hat{m}_k\right)^2$$

The residual variance[2] from *WLS* estimation of (3), unadjusted for degrees of freedom, is used as an estimate of the variance of the error terms due to omitted firm variables, $\hat{\sigma}_e^2$. Hence an estimate of the *omega* ratio is

$$\hat{K} = \hat{\sigma}_m^2/(\hat{\sigma}_m^2 + \hat{\sigma}_e^2)$$

This estimate of the *omega* ratio along with D_{ik} is used to calculate the estimated variance-covariance matrix, thus completing the first stage.

III. Illustration

The sample of firms that we use to illustrate the procedure is described fully in a paper by Vernon and R. E. M. Nourse. Briefly the sample consists of 57 large manufacturing firms engaged in selling food, beer, liquor and wine, tobacco, soaps and detergents, household supplies, and toiletries. The data cover the period 1963–69 and were obtained from the usual trade and governmental sources. Industry variables were constructed at the 4-digit SIC level of aggregation. The variables are defined below.

Dependent

R = the firm's net income divided by shareholders' equity, averaged over the years 1963–68

Independent

D_{ik} = the share of the ith firm's sales sold in market k

DC = a dummy variable that equals unity if the weighted[3] average concentration ratio of the firm's product markets is greater than 50, and zero otherwise (1966 4-firm concentration ratios).

CAS = the 1969 advertising/sales ratio of the firm.

AS = the weighted average industry advertising/sales ratio of the firm's product markets

LS = the reciprocal of the logarithm of 1968 total assets of the firm

We examined regressions for two specifications. In model A, the company advertising/sales ratio, CAS, is used as an independent variable along with a market concentration and a firm size variable; in model B, CAS is replaced by AS, the weighted industry advertising/sales ratio.[4]

We report our estimation results for stage one only summarily. Many variations of the stage one procedure were attempted.[5] In all cases the *omega* ratio estimates for models

fied firm is a linear combination of market intercepts. Therefore, in equation (3) each firm has a weighted dummy variable for each market in which it participates, the weights being the share of firm sales in that market. Because we have 40 markets and only 57 observations, the use of a dummy variable for each market would leave us with too few degrees of freedom. A reasonable compromise seemed to be to aggregate the 40 four-digit markets into 11 three-digit markets. Hence, we used 11 dummy variables rather than 40.

[2] Of course, the residual variance must be adjusted to correct for the weighting procedure.

[3] The weights used are the estimated shares of the firm's total sales in each of the 40 4-digit SIC industries covered.

[4] The two advertising variables would appear to raise some ambiguities with respect to the I-H scheme of classifying all variables as either firm or market related. For example, *AS* might be considered to be a bit of both.

[5] Among the variations that we tried were: *WLS* with the weighted dummies representing markets (the weights were $1/\sqrt{H_i}$), and *CLS* with both weighted and unweighted dummies. Imel has pointed out that *DC* is incorrectly constructed *if* one wishes to test the hypothesis that firms earn higher profits in markets where concentration exceeds 50 percent. Rather than defining the zero-one dummy variable on *CR* (as we have done), one should first define the dummies for each market and then construct the firm's variable (*DW*) as a weighted average of the market dummies. We therefore estimated our regressions substituting *DW* for *DC* and found only slight differences in the results. Given the simple correlation coefficient between *DW* and *DC* of .95, this is not unexpected. We should note, however, that the *DW* variable did tend to yield slightly higher t-values. For example, the *DW* t-value in the regression of the profit rate on *DW*, *AS*, and *CAS* was 1.62 as compared with 1.5 for *DC*.

TABLE 1—MULTIPLE REGRESSION EQUATIONS EXPLAINING FIRM PROFIT RATES USING CLASSICAL LEAST SQUARES (*CLS*) AND USING GENERALIZED LEAST SQUARES (*GLS*) WITH FOUR ALTERNATIVE SPECIFICATIONS OF THE VARIANCE-COVARIANCE MATRIX OF THE ERROR TERM

Estimation Procedure	Assumed Value of $\hat{K}$	Intercept	*DC*	*AS*	*CAS*	*LS*	R^2	Equation Number
CLS		.141	.009 (.44)		0.806 (4.27)[a]	−.260 (−1.50)	.28	(1)
GLS	0.0	.124	.019 (1.13)		0.790 (4.45)[a]	−.210 (−1.33)	.42	(2)
GLS	0.23	.138	.022 (1.09)		0.703 (3.78)[a]	−.268 (−1.69)	.38	(3)
GLS	0.5	.153	.019 (.83)		0.698 (3.68)[a]	−.337 (−2.06)[a]	.36	(4)
GLS	0.8	.181	.113 (.37)		0.722 (3.73)[a]	−.474 (−2.71)[a]	.32	(5)
CLS		.105	.028 (1.50)	0.833 (5.01)[a]		−.178 (−1.08)	.34	(6)
GLS	0.0	.111	.026 (1.58)	0.838 (4.96)[a]		−.185 (−1.21)	.46	(7)
GLS	0.21	.121	.023 (1.13)	0.823 (3.42)[a]		−.211 (−1.31)	.35	(8)
GLS	0.5	.132	.018 (.73)	0.883 (2.61)[a]		−.255 (−1.49)	.29	(9)
GLS	0.8	.145	.011 (.32)	1.091 (2.04)[a]		−.335 (−1.82)	.20	(10)

Note: The numbers in parentheses are *t*-ratios; the number of observations is 57.
[a] Indicates significance at the .025 level for a one-tail *t*-test.

A and *B* lay in the range of 0.173 to 0.247. For example, one set of estimates for model *A* yielded $\hat{\sigma}_m^2 = .0012$, $\hat{\sigma}_e^2 = .0043$, and $\hat{K} = 0.217$. The corresponding estimates for model B were $\hat{\sigma}_m^2 = .0012$, $\hat{\sigma}_e^2 = .0045$, and $\hat{K} = 0.207$. For use in stage two, we selected the following representative values of $\hat{K}$:

	$\hat{K}$ (the *omega* ratio)
Model *A*	0.23
Model *B*	0.21

To permit comparison of our results with those of I-H, we estimated models *A* and *B* using classical least squares (*CLS*) and using *GLS* conditional on the following values of the *omega* ratio: 0.00, 0.23 or 0.21, 0.50, and 0.80. These equations are presented in Table 1.

For every equation the signs of our estimated coefficients agree with a priori expectations. As I-H found, the significance of the concentration ratio and the advertising variables tend to fall as the *omega* ratio increases. Contrary to the I-H findings, the advertising/sales ratio remained significantly positive at a .025 level even when the *omega* ratio was set at 0.80. It is interesting to note that the *GLS* results for $K = 0.00$ (equivalent to *WLS*) and for $K = 0.23$ (or 0.21) are so similar that no qualitatively different conclusions can be drawn.

In *GLS* estimation of the I-H model, *ceteris paribus* the larger the *omega* ratio the

more weight is given to the off-diagonal elements of the variance-covariance matrix. That is, the more important are the omitted market variables relative to the omitted firm variables (i.e., the bigger is K), the more we need to take into account the interdependence among error components for firms selling in the same markets. A large K emphasizes this interdependence at the expense of differences among firms' levels of diversification and conversely. In market structure-performance studies prior to that of I-H, the usual practice was to use weighted least squares (i.e., $K=0$) and to neglect the off-diagonal elements of the variance-covariance matrix altogether. Under the I-H specification, this problem would be misleading if in fact K were large. The evidence we have put forward in this note suggests that although K is not zero, nonetheless it is not large enough to make the *GLS* results very different from those for *WLS*. However, we should emphasize that our results are specific to our sample of data.

A final caveat on the use of the I-H model is in order. In order to use their model, it is necessary to have an unambiguous classification of the independent (and omitted) variables as being either firm-related or market-related. This does not always appear to be possible. We have already noted the doubtful classification of the advertising variables in this respect. Another example is a measure of firm size: the *average* size of firms in a market might be viewed as a market-related variable and deviations from the average as firm-related, or the measure might be viewed solely as a firm-related variable (following I-H).

Finally, we do consider the I-H work to be a useful contribution to the growing literature on market structure-performance studies where observations are on firms rather than industries. If for no other reason, I-H are to be commended for highlighting the rather severe assumption that the profit equation of a diversified firm is a simple *linear* combination of profit equations of the firm's divisions.

REFERENCES

M. Hall and L. W. Weiss, "Firm Size and Profitability," *Rev. Econ. Statist.*, Aug. 1967, *49*, 319–31.

B. Imel and P. Helmberger, "Estimation of Structure-Profit Relationships with Application to the Food Processing Sector," *Amer. Econ. Rev.*, Sept. 1971, *61*, 614–27.

M. Nerlove, "Further Evidence on the Estimation of Dynamic Economic Relations from a Time Series of Cross Sections," *Econometrica*, Mar. 1971, *39*, 359–82.

J. M. Vernon, *Market Structure and Industrial Performance: A Review of Statistical Findings*, Boston 1972.

——— **and R. E. M. Nourse**, "The Effect of Market Structure on Profit Rates of Advertising Intensive Firms," *J. Ind. Econ.*, forthcoming.

Federal Trade Commission, *Economic Report on the Influence of Market Structure on the Profit Performance of Food Manufacturing Companies*, September 1969.

———, *New York Times*, Jan. 25, 1971, p. 1.

Estimation of Structure-Profit Relationships: Reply

By Blake Imel and Peter Helmberger*

Recent work has focused attention on the need for more careful specification of the variance-covariance matrix for the error term in the estimation of structure-profit rate relationships. It has been shown that where profit rate data for individual companies are analyzed, the matrix will likely display both heteroskedasticity and autocorrelation.[1] The *omega* ratio, relevant to autocorrelation, is defined as the ratio of the variance of the error component attributable to the omission of market related variables, σ_m^2, to the total variance of the error term, σ_u^2. This ratio appears in every element of the *omega* matrix, defined as $1/\sigma_u^2$ times the variance-covariance matrix, where two firms in the sample represent the same industry. In their comment, John Vernon and Marjorie McElroy estimate that the *omega* ratio is in the neighborhood of .22 and suggest, subject to the need for analysis of additional samples, that neglecting the problem of autocorrelation has rather negligible effects on parameter estimates. In what follows, we draw attention to possible shortcomings of their analysis and argue that the problem of autocorrelation should not be dismissed on the basis of their findings.

Vernon and McElroy propose adapting a two-stage technique developed by Marc Nerlove to estimate the *omego* ratio. They note that the implementation of this technique would entail including a dummy variable for all 40 markets, save one, represented by the 57 firms in their sample. Arguing that this would severely tax the degrees of freedom, they aggregate the 40 sample markets into 11 SIC three-digit groups and hypothesize a dummy variable for each group. We feel that this procedure is not only inappropriate in light of their objectives but indeed, raises a whole host of new problems.

To see this, consider a sample of specialized firms where profit rates are hypothesized to be a function of variables related to firms, markets, and sectors. The total error could then be broken up into component parts as follows:

$$u_{ijt} = e_{ijt} + m_{jt} + g_t \tag{1}$$

where u_{ijt} equals the total error for the ith firm in the jth market in the tth sector and the components e_{ijt}, m_{jt}, and g_t are attributable to the omission of variables related to firms, markets, and sectors, respectively. In our earlier paper, it was assumed that sector related variables and sector related error do not exist; the t subscripts were not required. On this assumption, Vernon and McElroy have included irrelevant variables—the sector dummies—and excluded relevant variables—the market dummies. (This conclusion is hardly altered by allowing for diversification because the equations for diversified firms are mere aggregates of equations for the component "specialized firms.") Theoretically, the Vernon-McElroy estimates of m_j are biased downward toward zero. Indeed, if there were zero correlation between each sector dummy and excluded relevant variables, the estimates of m_j would not be expected to differ significantly from zero. The estimate of the *omega* ratio is biased downward for two reasons. The estimate of the numerator is biased downward because of the downward bias in the estimates of m_j. The denominator is biased upward due to the omission of the relevant market dummies. Clearly, if assuming g_t is nonexistent is valid, the Vernon-McElroy estimate of the *omega* ratio is open to question.

The other alternative, of course, is that relevant sector variables and sector error may in fact exist. What then? For specialized firms, it can be shown that the covariance of the error terms for the two firms in the

* U. S. Department of Agriculture and University of Wisconsin, Madison, respectively.

[1] See our 1971 article. For a discussion of the implications of error component analysis for aggregative approaches, see Imel, Michael Behr, and Helmberger, pp. 77–84.

same market, and therefore in the same sector equals the variance of m_j, σ_m^2, plus the variance of g_t, σ_g^2. The appropriate element of the *omega* matrix would consist of the *omega* ratio plus the ratio of σ_g^2 to σ_u^2. For two firms in the same sector but in different markets, the appropriate element of the *omega* matrix simply equals the ratio of σ_g^2 to σ_u^2. Because of the many patterns of diversification among markets and/or sectors which are conceptually possible, statistical analysis of the error components becomes very complex if the sample contains diversified firms. Moreover, we have some reservations regarding the operationality of the statistical theorems that can be derived. Instead of estimating one variance ratio in the *omega* matrix, one would need estimates of two variance ratios. Application of the Nerlove technique would apparently involve including sector dummies as well as market dummies. It does not appear to us that Vernon and McElroy could have skirted the problem of insufficient degrees of freedom through postulating a sector error component if indeed that was their intent.

Our third and final concern is that the more-or-less straightforward application of Nerlove's suggested technique for the particular problem at hand may not be completely appropriate. As described by Nerlove the technique involves estimating the off-diagonal parameters of the variance covariance matrix when there are T observations on each of N individuals (markets in this case). The variance among the estimated coefficients of the N dummy variables is taken as an estimate of the variance of the "individual related" terms (market related in this case). In industrial organization research, only by coincidence would each market be represented by the same number of sample firms. This raises some substantive questions about the equal weighting each estimated dummy coefficient would receive in computing the variance via the Nerlove technique. An example might clarify this point.

Consider a representative sample of firms from a hypothetical world where all firms are completely specialized. A regression of profit rates on explanatory variables is estimated with a dummy variable included for each market in the sample. In estimating σ_m^2, Nerlove's technique gives each estimated market related term (dummy coefficient) equal weight. Yet, in the sample variance of m_j across firms, some of the market related terms might well be represented many times relative to others. Equal weighting would appear to be a questionable procedure. It is possible, using equal weights, that the estimated market related variance, σ_m^2, will exceed the total variance of the dependent variable in either the sample or population.[2]

In conclusion, while we are sympathetic to the need for estimates of the *omega* ratios in cross-section studies, we must unfortunately report some dissatisfaction with the estimates presented by Vernon and McElroy for their particular study. Though they are to be commended for calling attention to the possibility of using Nerlove's technique to estimate *omega* ratios, their application of this technique has been found wanting in a number of respects. In our view it would be unacceptable in future regressions on company profit rates to assume that the problem of autocorrelation can be neglected on the basis of the Vernon-McElroy findings.

REFERENCES

B. Imel and P. Helmberger, "Estimation of Structure-Profit Relationships with Application to the Food Processing Sector," *Amer. Econ. Rev.*, Sept. 1971, *61*, 614–27.

B. Imel, M. Behr, and P. Helmberger, *Market Structure and Performance*, Lexington 1972.

M. Nerlove, "Further Evidence on the Estimation of Dynamic Economic Relations from a Time Series of Cross Sections," *Econometrica*, Mar. 1971, *39*, 359–82.

[2] Another way of viewing the problem is to observe that giving each $\hat{m}_j$ equal weight in the estimation of $\hat{\sigma}_m^2$ is to compute the variance of m_j across markets only. In a regression on firm data however, the variance of m_j is among firms which might be quite unevenly distributed among markets. The difference is in terms of what one is summing over in the variance expression.

Relative Shares and Elasticities Simplified: Comment

By PAUL A. SAMUELSON*

The taxonomy of changing factor shares in a competitive industry's output, as factor supplies (or prices) shift, has, again, been ably discussed.[1] These results can be included under the following terse exposition.

I

Denote the neoclassical (homogeneous-first-degree, concave) production function as $q=f(x_1, \ldots, x_n)$. Then write $X_i = log_e x_i = ln\ x_i$ and

$$(1)\quad ln\, f(x_1, \ldots, x_n) = ln\, f(e^{X_1}, \ldots, e^{X_n}) = \alpha(X_1, \ldots X_n)$$

The relative share of the ith factor,

$$p_i x_i \Big/ \sum_1^n p_j x_j = \alpha_i$$

is then given by

$$(2)\quad \alpha_i = \partial\alpha(X_1, \ldots, X_n)/\partial X_i = \alpha_i(X_1, \ldots, X_n)$$

where subscripts on a function denote partial differentiation, as in $\partial\alpha/\partial X_i = \alpha_i, \partial^2\alpha/\partial X_i \partial X_j = \alpha_{ij}$, etc.

By definition, total relative shares sum to unity; hence

$$(3)\quad \partial\left(\sum_1^n \alpha_i\right) \Big/ \partial X_j = 0 = \sum_1^n \alpha_{ij}(X_1, \ldots, X_n)$$

Effects on α_i, from a relative change in factor i alone are given simply by the sign of α_{ii}, or equivalently of $\alpha_{ii}/\alpha_i\alpha_i$, and $\alpha_{ii}/\alpha_i(1-\alpha_i)$ (the latter being $1-\sigma_i^{-1} = 1 + [x_i f_{ii}/(1-\alpha_i) f_i]$, the "Samuelson" criterion). Effects on α_i from a sole change in factor j are given by α_{ij} or $\alpha_{ij}/\alpha_i\alpha_j$ (which are closely related to the "partial elasticities" concepts, $ff_{ij}/f_i f_j$, of Hicks, Allen, McFadden, Sato-Koizumi, and many others).

To determine how the share of some subset of factors, say $\alpha_1 + \ldots + \alpha_r$, $r<n$, changes relative to some other subset's, $\alpha_{r+1} + \ldots + \alpha_m$, $m<n$, when some factor x_j alone changes, we need only calculate the sign of

$$(4)\quad \partial\left(ln \sum_1^r \alpha_i - ln \sum_{r+1}^m \alpha_i\right) \Big/ \partial X_j = \left(\sum_1^r \alpha_i\right)^{-1}\left(\sum_1^r \alpha_{ij}\right) - \left(\sum_{r+1}^m \alpha_i\right)^{-1}\left(\sum_{r+1}^m \alpha_{ij}\right)$$

Thus, if $r=2$, $m=3=j$, as in Sato-Koizumi's example of a change in brown labor's supply and the effects of this upon its share relative to that of white and black laborers' combined, (4) becomes $(\alpha_1+\alpha_2)^{-1}(\alpha_{13}+\alpha_{23}) - \alpha_{33}^{-1}\alpha_{33}$. If $m=n=3$, (4) becomes $-\alpha_{33}[(1-\alpha_3)^{-1} + (\alpha_3)^{-1}]$.

II

This completes all there is to say.[2]

Mention, however, should be made of the obvious duality relations that hold for the minimum-unit cost function in terms of factor prices, $c(p_1, \ldots, p_n)$, which is dual to the production function and has all its qualitative properties. Now we work with

$$(5)\quad \alpha[Y_1, \ldots, Y_n] = ln\, c(e^{Y_1}, \ldots, e^{Y_n})$$
$$\alpha^i[Y_1, \ldots, Y_n] = \partial\alpha[Y_1, \ldots, Y_n]/\partial Y_i$$

where now brackets (rather than parentheses) differentiate between factor prices and factor quantities, and superscripts (rather than subscripts) indicate differentiation with

* Institute professor, Massachusetts Institute of Technology. I owe thanks to the National Science Foundation.

[1] See Ryuzo Sato and Tetsunori Koizumi. This gives references to Hicks, Robinson, McFadden, Samuelson, Sato-Koizumi, and others.

[2] I have skipped the Sato-Koizumi discussion of "direct and shadow elasticities," dealing with a change in "x_i/x_j," since, when $n>2$, such Robinsonian changes are ill-defined

respect to factor prices (rather than factor quantities).

Now, if we ask the effects on factor shares of changes in one of many factor prices (instead of factor quantities), all our previous formulas can immediately have their duals written, merely by changing subscripts to superscripts! Thus, instead of α_{ii} we write α^{ii}, etc.[3]

[3] The informed reader can verify the following symmetric relations between the "dual" functions:

$$\alpha[Y_1, \dots, Y_n] + \alpha(X_1, \dots, X_n) \geq 0 \tag{6}$$

for $Y_i = ln(p_i / \sum p_j x_j)$ and with (6)'s equality sign holding if, and only if,

$$\begin{aligned} Y_i &= -X_i + ln\, \alpha_i(X_1, \dots, X_n) \\ X_i &= -Y_i + ln\, \alpha_i[Y_1, \dots, Y_n] \end{aligned} \tag{7}$$

In conclusion, and with no implication of criticizing earlier writers (including myself), this terse exposition may usefully point up the fact that most elasticity discussions are empty of substance: for the most part, whether they are partial elasticities or total elasticities, they merely give names to effects that need to be measured—except for their occasional function of relating some of the different needed measurements to each other.

REFERENCES

R. Sato and T. Koizumi, "The Production Function and the Theory of Relative Shares," *Amer. Econ. Rev.*, June 1973, *63*, 484–89.

Relative Shares and Elasticities Simplified: Reply

By RYUZO SATO AND TETSUNORI KOIZUMI*

We are gratified that mention of our paper on "Relative Shares" has been made in the preceding note by Paul Samuelson. Consistent with his past contributions, Samuelson has provided a simplified treatment of the various concepts of elasticity of substitution. We would like to take this occasion to make some further observations.

With regard to Samuelson's statement about the usefulness of the elasticity concepts, first, it is hard to imagine that Samuelson would go so far as to belittle the theoretical contributions of all the studies done in such areas as growth, production functions, and technical progress which have greatly benefited by the use of the elasticity of substitution concept. In these studies the elasticity of substitution concept was not used in a trivial manner. Secondly, and this is more relevant to the type of problems we address, there is the matter of substitutes and complements which is an integral aspect of the multifactor analysis. In problems dealing with multifactor production, dual partial elasticity concepts developed by Allen, Hicks, Sato-Koizumi, etc.[1] serve several useful functions. Besides providing a taxonomy of "effects that need to be measured," these various concepts provide us with alternative frameworks of conceiving of substantive problems.

Although it can be shown that the partial concepts are related to some of the other elasticity concepts and hence need not be glorified by fancy names, there do exist particular problems where it is crucial to understand how the "total" effect is divided into its "component" effects. For example, in the case where we are dealing with derived demand with some factors (subsitutes or complements) limited in supply, it is important to know exactly how the existence of substitutability or complementarity among factors works for or against a particular factor share.

* Professor of economics, Brown University and associate professor of economics, Ohio State University, respectively.

[1] See the references cited in our original article.

REFERENCES

P. A. Samuelson, "Relative Shares and Elasticities Simplified: Comment," *Amer. Econ. Rev.*, Sept. 1973, *63*, 770–71.

R. Sato and T. Koizumi, "The Production Function and the Theory of Relative Shares," *Amer. Econ. Rev.*, June 1973, *63*, 484–89.

NOTES

EIGHTY-SIXTH ANNUAL MEETING OF THE AMERICAN ECONOMIC ASSOCIATION

New York, New York, December 27–30, 1973

Preliminary Announcement of the Program

Thursday, December 27, 1973

2:00 P.M. EXECUTIVE COMMITTEE MEETING

Friday, December 28, 1973

8:30 A.M. PERSPECTIVES ON TWENTIETH CENTURY ECONOMIC THEORY (Joint Session with the History of Economics Society)
Chairman: ROBERT V. EAGLY, State University of New York, Binghamton
Papers: JOSEPH J. SPENGLER, Duke University and University of North Carolina
Was 1922–72 A Golden Age in the History of Economics?
SIR JOHN HICKS, Oxford University
Capital Controversy: Ancient and Modern
Discussants: CARL G. UHR, University of California, Riverside
ROBERT W. CLOWER, University of California, Los Angeles

8:30 A.M. WHAT SHOULD AMERICA'S HOUSING POLICIES BE?* (Joint Session with the American Finance Association)
Chairman: ANTHONY DOWNS, Real Estate Research Corporation
Papers: JOHN KAIN, Harvard University
What Should America's Housing Policies Be? I
FRANK DE LEEUW, The Urban Institute
What Should America's Housing Policies Be? II
Discussants: JAMES L. SWEENEY, Stanford University
MAHLON R. STRASZHEIM, University of Maryland
BENJAMIN CHINITZ, Brown University

8:30 A.M. THE FLOW OF FUNDS AND INTEREST RATES I: U.S. FINANCIAL MODELS* (Joint Session with the American Finance Association and the American Statistical Association)
Chairman: LAWRENCE S. RITTER, New York University
Papers: JAMES S. DUESENBERRY, Harvard University AND BARRY BOSWORTH, The Brookings Institution*
A Flow-of-Funds Model of the U.S. Financial System
JAMES L. PIERCE AND THOMAS THOMSON, Federal Reserve Board
Short-Term Financial Models at the Federal Reserve Board
Discussants: STEPHEN M. GOLDFELD, Princeton University
ALLAN H. MELTZER, Carnegie-Mellon University
DANIEL H. BRILL, Commercial Credit Company

8:30 A.M. A CRITIQUE OF GRANTS ECONOMICS* (Joint Session with the Association for the Study of the Grants Economy)
Chairman: EVSEY DOMAR, Massachusetts Institute of Technology
Papers: KENNETH E. BOULDING, University of Colorado
Micro-Grants-Economics
MARTIN PFAFF, Wayne State University and University of Augsburg
Macro-Grants-Economics
Discussants: JAMES N. MORGAN, University of Michigan
WARREN SAMUELS, Michigan State University

* Not to appear in the *Papers and Proceedings*

10:30 A.M. THE ECONOMICS OF THE FIRST AMENDMENT
Chairman: LEONARD SILK, The New York Times
Papers: RONALD H. COASE, University of Chicago
The Market for Goods and the Market for Ideas
WILLIAM F. BAXTER, Stanford University Law School
On Regulation, Efficiency, and Diversity in Communications Media
Discussants: ROGER NOLL, The Brookings Institution
BRUCE OWEN, Stanford University
ALICE VANDERMUELEN, University of California, Los Angeles

10:30 A.M. THE CONSEQUENCES OF DEMOGRAPHIC CHANGES FOR ECONOMIC POLICY
Chairman: JOSEPH J. SPENGLER, Duke University
Papers: RONALD G. RIDKER, Resources for the Future
Resource and Amenity Implications of Changes in Population Growth Rates
ALAN SWEEZY AND AARON OWENS, California Institute of Technology
The Impact of Changes in Population Growth on Employment
ALLEN C. KELLEY, Duke University
The Role of Population in Models of Economic Growth
Discussants: ANSLEY J. COALE, Princeton University
JOHN R. MEYER, National Bureau of Economic Research and Yale University

10:30 A.M. ENDOWMENT INCOME, CAPITAL GAINS AND INFLATION ACCOUNTING (Joint Session with the American Finance Association)
Chairman: W. ALLEN WALLIS, University of Rochester
Papers: JAMES TOBIN, Yale University
What is Permanent Endowment Income?
JAMES M. LITVACK, BURTON MALKIEL, AND RICHARD E. QUANDT, Princeton University
A Plan for the Definition of Endowment Income
DONALD A. NICHOLS, University of Wisconsin
The Investment Income Formula of the American Economic Association
Discussants: ROBERT EISNER, Northwestern University
J. PETER WILLIAMSON, Dartmouth College
STANLEY W. BLACK, Vanderbilt University

10:30 A.M. WAGE-PRICE CONTROLS I: TWO YEARS OF WAGE-PRICE CONTROLS: WHAT HAVE WE LEARNED? (Joint Session with the American Statistical Association)
Chairman: GEORGE L. PERRY, The Brookings Institution
Papers: BARRY BOSWORTH, The Brookings Institution
Why Phase III Failed
MARY HAMILTON, Loyola University
Price Controls in 1973
ROBERT LANZILLOTTI AND BLAINE ROBERTS, University of Florida
An Inside View of Controls
J. B. MITCHELL, University of California, Los Angeles, AND ARNOLD WEBER, Carnegie-Mellon University
Wages and the Pay Board
Discussants: PHILIP CAGAN, Columbia University
ROBERT SOLOW, Massachusetts Institute of Technology

10:30 A.M. PRODUCTION AND THE AGGREGATE DEMAND FUNCTION FOR LABOR
Chairman: LAWRENCE R. KLEIN, University of Pennsylvania
Papers: NEIL WALLACE AND THOMAS SARGENT, University of Minnesota and The Federal Reserve Bank of Minneapolis
The Elasticity of Substitution and the Cyclical Behavior of Productivity, Average Earnings, and Labor Share
SHERWIN ROSEN, Harvard University and M. I. NADIRI, National Bureau of Economic Research
A Disequilibrium Model of Demand for Factors of Production
ROSS S. PRESTON, University of Pennsylvania
A System of Industry Disaggregate Labor Requirements Equations in an Input-Output Framework
Discussants: ROBERT HALL, Massachusetts Institute of Technology
MICHAEL D. MCCARTHY, Case Western Reserve University
ANNE CARTER, Brandeis University

12:30 P.M. JOINT LUNCHEON* (With the American Finance Association and the American Statistical Association)
Chairman: WALTER W. HELLER, University of Minnesota
Speaker: GEORGE P. SHULTZ, Secretary of the Treasury

2:00 P.M. MONETARY POLICY AND SECTORAL CREDIT FLOWS
Chairman: JOHN KAREKEN, University of Minnesota
Papers: JAMES S. DUESENBERRY, Harvard University
Analysis of the Issues
ANDREW F. BRIMMER, Federal Reserve Board
Alternative Strategies to Influence Credit Flows
Discussants: CRAIG SWAN, University of Minnesota
JOHN M. CULBERTSON, University of Wisconsin
WILLIAM POOLE, Federal Reserve Bank of Boston

2:00 P.M. WAGE-PRICE CONTROLS II: APPRAISAL OF U.S. WAGE AND PRICE CONTROL PROGRAMS* (Joint Session with the American Statistical Association)
Chairman: MORRIS COHEN, Long Island University
Panelists: ROBERT J. EGGERT, RCA Corporation
A View from Industry
GEORGE W. MCKINNEY, Irving Trust Company
A View from the Financial Sector
FRANK W. SCHIFF, Committee for Economic Development
A View from Washington
ABBA P. LERNER, Queens College, City University of New York
A View from Academe

2:00 P.M. INTERNATIONAL TRADE AND FINANCE I: THE POLITICAL ECONOMY OF MULTINATIONAL ENTERPRISES (Joint Session with the Academy of International Business, the American Finance Association, and the American Statistical Association)
Chairman: TIBOR SCITOVSKY, Stanford University
Papers: GARY C. HUFBAUER, University of New Mexico
Problems and Policies in the International Adjustment Process
GEORGE HILDEBRAND, Cornell University
Problems and Policies Affecting Labor's Interests
RAYMOND VERNON, Harvard University
Problems and Policies Respecting Competition and Monopoly Power
Discussants: C. FRED BERGSTEN, The Brookings Institution
PEGGY MUSGRAVE, Northeastern University
ROBERT B. STOBAUGH, Harvard University
THOMAS HORST, The Brookings Institution
JOHN J. MCGOWAN, Yale University

2:30 P.M. WORLD PREJUDICE AND SEX DISCRIMINATION: Findings of The American Economic Association Committee on the Status of Women in the Economics Profession
Chairman: CAROLINE SHAW BELL, Wellesley College
Committee Members: WALTER ADAMS, Michigan State University
KENNETH J. ARROW, Harvard University
FRANCINE BLAU, Trinity College
MARTHA BLAXALL, National Academy of Sciences
KENNETH BOULDING, University of Colorado
JOHN KENNETH GALBRAITH, Harvard University
COLLETTE MOSER, Michigan State University
BARBARA REAGAN, Southern Methodist University
MYRA STROEBER, Stanford University
PHYLLIS WALLACE, Sloan School, Massachusetts Institute of Technology
Panelists: ELIZABETH CLAYTON, University of Missouri
DAVID GORDON, New School for Social Research
(third to be announced)

* Not to appear in the *Papers and Proceedings*

2:30 P.M. MAJOR ECONOMIC POLICY PROBLEMS OF THE 1970s
Chairman: R. A. GORDON, University of California, Berkeley
Panelists: SAMUEL BOWLES, University of Massachusetts
ROBERT L. HEILBRONER, New School for Social Research
KERMIT GORDON, The Brookings Institution
HENDRIK S. HOUTHAKKER, Harvard University
GEOFFREY H. MOORE, National Bureau of Economic Research

2:30 P.M. THE PROPERTY TAX: PROGRESSIVE OR REGRESSIVE?
Chairman: JOSEPH A. PECHMAN, The Brookings Institution
Papers: HENRY J. AARON, The Brookings Institution
Views on Property Taxation: One
RICHARD A. MUSGRAVE, Harvard University
Views on Property Taxation: Two
Discussants: HARVEY E. BRAZER, University of Michigan
ANN F. FRIEDLAENDER, Boston College
DICK NETZER, New York University
GEORGE E. PETERSON, Urbana Institute
EARL R. ROLPH, University of California, Berkeley

3:00 P.M. THE GREAT SOCIETY VERSUS THE NEW FEDERALISM: AN ECONOMIC PERSPECTIVE
Chairman: Congressman PAUL SARBANES
Papers: CHARLES L. SCHULTZE, The Brookings Institution
Sorting Out the Arguments: An Economist's Criteria
LESTER THUROW, Massachusetts Institute of Technology
Cash Versus In-Kind Transfers: Is Economics Myopic About the Tradeoff?
Discussants: KENNETH J. ARROW, Harvard University
(Others to be announced)

8:00 P.M. RICHARD T. ELY LECTURE
Chairman: WALTER W. HELLER, University of Minnesota
Speaker: ROBERT M. SOLOW, Massachusetts Institute of Technology

Saturday, December 29, 1973

8:30 A.M. ALLOCATING RESOURCES FOR POLICY RESEARCH
Chairman: JAMES N. MORGAN, University of Michigan
Papers: F. THOMAS JUSTER, University of Michigan and Institute for Social Research
How Could Survey Research Be More Useful?
ALICE M. RIVLIN, The Brookings Institution
How Could Experiments Be More Useful?
Discussants: BARBARA R. BERGMANN, University of Maryland
MICHAEL H. MOSKOW, U.S. Department of Housing and Urban Development

8:30 A.M. INTERNATIONAL ENERGY SUPPLY
Chairman: SAM H. SCHURR, Resources for the Future
Papers: M. A. ADELMAN, Massachusetts Institute of Technology
Crude Oil Availability and Price Outside the United States
JAMES W. MCKIE, University of Texas
The Political Economy of International Petroleum
Discussants: RAYMOND F. MIKESELL, University of Oregon
THOMAS R. STAUFFER, Harvard University

8:30 A.M. ECONOMICS IN CHINA: THE PROSPECTS FOR SCIENTIFIC INTERCHANGE* (Joint Session with the Association for Comparative Economic Studies)
Chairman: LLOYD G. REYNOLDS, Yale University
Panelists: CARLOS DIAZ ALEJANDRO, Yale University
JOHN GURLEY, Stanford University
DWIGHT PERKINS, Harvard University

* Not to appear in the *Papers and Proceedings*

8:30 A.M. THE FLOW OF FUNDS AND INTEREST RATES II: FINANCIAL ANALYSIS THROUGH THE FLOW OF FUNDS* (Joint Session with the American Statistical Association and the American Finance Association)
Chairman: ALAN GREENSPAN, Townsend, Greenspan, and Company
Papers: DIMITRI BALATSOS, Manufacturers Hanover Trust Company
Homogenizing the Concepts of Financial Market Analysis
PATRIC HENDERSHOTT, Purdue University
Progress Report on a Macro Model of the Financial System
STEPHEN P. TAYLOR, Federal Reserve Board
State of the Statistical Art in Flow-of-Funds Accounting
Discussant: JAMES MCKEON, Salomon Brothers

8:30 A.M. LEXICOGRAPHIC ORDERING AND ITS APPLICATION* (Joint Session with the Econometric Society)
Chairman: NICHOLAS GEORGESCU-ROEGEN, Vanderbilt University
Papers: RICHARD H. DAY AND STEPHEN M. ROBINSON, University of Wisconsin
Economic Decisions with L** Utility
HOWARD KUNREUTHER, University of Pennsylvania
Protection Against Natural Hazards: A Lexicographic Approach
WILLIAM T. TERRELL, Wichita State University
Lexicographic Choice: Pure and Applied Economic Theory
Discussant: JOHN S. CHIPMAN, University of Minnesota

10:30 A.M. THE FIGHT AGAINST POVERTY: CONFLICTING OBJECTIVES AND ALTERNATIVE STRATEGIES FOR INCOME MAINTENANCE (Joint Session with the Association for the Study of the Grants Economy)
Chairman: ANDREW F. BRIMMER, Federal Reserve Board
Papers: ALAIR TOWNSEND AND ROBERT LERMAN, Joint Economic Committee, U.S. Congress
Conflicting Objectives in Income Maintenance Programs: Is There an Easy Way Out?
ROBERT H. HAVEMAN AND IRWIN GARFINKEL, University of Wisconsin
Earnings Capacity and the Anti-Poverty Effectiveness of Income Maintenance Alternatives
Discussants: WORTH BATEMAN, Urban Institute
J. R. DOMINGUEZ, University of California, Los Angeles
PHYLLIS WALLACE, Massachusetts Institute of Technology
VINCENTE T. XIMENES, former Commissioner, Equal Employment Opportunity Commission

10:30 A.M. THE THEORY OF POLICY (Joint Session with the Econometric Society)
Chairman: BENT HANSEN, University of California, Berkeley
Papers: WALTER P. HELLER AND KARL SHELL, University of Pennsylvania
On the Theory of Optimal Taxation with Costly Administration
THOMAS MUENCH AND NEIL WALLACE, University of Minnesota and the Federal Reserve Bank of Minneapolis
On Finding a Good Stabilization Policy: Models and Methods
Discussants: ANTHONY B. ATKINSON, University of Essex
EDMUND S. PHELPS, Columbia University

10:30 A.M. LESSONS FROM AND FOR THE CURRENT ECONOMIC EXPANSION
Chairman: ARTHUR M. OKUN, The Brookings Institution
Panelists: OTTO ECKSTEIN, Harvard University
WILLIAM J. FELLNER, American Enterprise Institute
PAUL W. MCCRACKEN, University of Michigan
PAUL A. SAMUELSON, Massachusetts Institute of Technology

10:30 A.M. INTERNATIONAL TRADE AND FINANCE II: U.S. COMPETITIVENESS IN WORLD MARKETS* (Joint Session with the American Statistical Association and the American Finance Association)
Chairman: ROBERT G. HAWKINS, New York University
Papers: J. DAVID RICHARDSON, University of Wisconsin
Estimating Demand and Supply Parameters for Imports without Import Prices: Methodology and Empirical Results
HELEN JUNZ AND BARBARA LOWREY, Federal Reserve Board
The Effect of Changes in Economic Activity on the U.S. Trade Balance in the Monitoring System for the Balance-of-Payments Process
RICHARD A. LABARGE, Louisiana State University
Foreign Penetration of the Automobile Market in the United States
Discussant: JAMES F. BURTLE, W. R. Grace & Company

* Not to appear in the *Papers and Proceedings*

10:30 A.M. WOMEN AT WORK: AT HOME AND IN THE MARKET (Joint Session with the American Statistical Association)
Chairman: MARINA V. N. WHITMAN, University of Pittsburgh
Papers: MICHAEL J. BOSKIN, Stanford University
The Effects of Government Taxes and Expenditures on Female Labor
ARLEEN LEIBOWITZ, National Bureau of Economic Research and Brown University
Production Within the Household
VICTOR R. FUCHS, City University of New York and National Bureau of Economic Research
Short-Run and Long-Run Prospects for Female Earnings
Discussants: MARY L. EYSENBACH, University of Utah
MICHAEL PIORE, Massachusetts Institute of Technology
ISABEL V. SAWHILL, Goucher College

1:30 P.M. CONCEPTS OF DISTRIBUTIONAL EQUITY (Joint Session with the Public Choice Society)
Chairman: CARL KAYSEN, Princeton University
Papers: JOHN RAWLS, Harvard University
Distributional Equity
DENNIS MUELLER, Cornell University
Defining the Polity for Distributional Norms
JAMES M. BUCHANAN AND WINSTON BUSH, Virginia Polytechnic Institute and State University
Positive Limits on Distributional Policy
Discussants: LESTER THUROW, Massachusetts Institute of Technology
ALVIN KLEVORICK, Yale University

2:00 P.M. ENERGY SUPPLY AND DEMAND IN THE UNITED STATES* (Joint Session with the American Agricultural Economics Association)
Chairman: DUANE CHAPMAN, Cornell University
Papers: STEPHEN L. MCDONALD, University of Texas
Incentive Policy and Supplies of Energy Sources
HARRY PERRY AND MILTON F. SEARL, Resources for the Future
Policies for Energy Research and Development
HENDRIK HOUTHAKKER, Harvard University AND PHILIP VERLEGER, Data Resources, Inc.
Demand Studies for Energy Resources
JOHN WILSON, Federal Power Commission
Supply-Demand Interactions: Environmental Protection, Price Policy, and Industrial Organization
Discussants: A. MYRICK FREEMAN, Bowdoin College
RICHARD L. GORDON, Pennsylvania State University
ROBERT KALTER, Cornell University
WALTER MEAD, University of California, Santa Barbara

2:30 P.M. ECONOMIC EDUCATION
Chairman: ROBERT L. HEILBRONER, New School for Social Research
Papers: KENNETH BOULDING AND ELISE BOULDING, University of Colorado
Introducing Freshmen to the Social Sciences
RENDIGS FELS, Vanderbilt University
Developing Independent Problem-Solving Ability in Economics
BERNARD BOOMS, Pennsylvania State University
Computer-Aided Instruction for Large Elementary Courses
Discussants: LESTER FETTER, Ocean County College
HENRY H. VILLARD, City University of New York
ELISABETH ALLISON, Harvard University

2:30 P.M. TRENDS, PROSPECTS AND POLICIES FOR EAST-WEST TRADE (Joint Session with the Association for Comparative Economic Studies)
Chairman: WALTER GALENSON, Cornell University
Papers: ALEXANDER ECKSTEIN, University of Michigan
Prospects and Policy for Sino-U.S. Trade
GREGORY GROSSMAN, University of California, Berkeley
Prospects and Policy for Soviet-U.S. Trade

* Not to appear in the *Papers and Proceedings*

EGON NEUBERGER, State University of New York, Stony Brook; ALAN A. BROWN, University of Windsor; and PAUL MARER, University of Indiana
Prospects and Policy for Trade Between Eastern Europe and the United States
Discussants: T. C. LIU, Cornell University
LAWRENCE J. BRAINARD, The Chase Manhattan Bank
GILBERT RODGERS, U.S. Department of Commerce

3:00 P.M. IS ECONOMIC THEORY WITH IT?
Chairman: ROBERT M. SOLOW, Massachusetts Institute of Technology
Panelists: HARRY G. JOHNSON, University of Chicago and London School of Economics
VERNON L. SMITH, California Institute of Technology
GARY S. BECKER, University of Chicago
TJALLING C. KOOPMANS, Yale University
JOHN KENNETH GALBRAITH, Harvard University

8:00 P.M. PRESIDENTIAL ADDRESS
Chairman: (To be announced)
Speaker: KENNETH J. ARROW, Harvard University

9:15 P.M. BUSINESS MEETING

Sunday, December 30, 1973

8:30 A.M. INTERCITY PASSENGER TRANSPORTATION: AILMENTS AND PRESCRIPTIONS (Joint Session with the Transportation and Public Utility Group)
Chairman: ROY J. SAMPSON, University of Oregon
Papers: HUGH S. NORTON, University of South Carolina
The Wheel: Should We Re-Invent It?
EDWIN P. PATTON, University of Tennessee
Where Goest the Pointless Arrow?
GEORGE C. EADS, George Washington University
Airline Capacity Limitation Controls: Private Vice Equals Public Virtue?
Discussant: JAMES C. NELSON, Washington State University

8:30 A.M. UNEMPLOYMENT IN A FULL EMPLOYMENT ECONOMY* (Joint Session with the American Statistical Association)
Chairman: CHARLES C. HOLT, Urban Institute
Papers: HYMAN KAITZ, U.S. Bureau of Labor Statistics
Unemployment Issues
RALPH E. SMITH, Urban Institute
The Discouraged Worker in a Full Employment Economy
HERMAN P. MILLER, Economic Consultant
Subemployment in the United States
Discussant: BELTON FLEISHER, Ohio State University

8:30 A.M. POSTWAR ECONOMIC GROWTH IN EASTERN EUROPE* (Joint Session with the Association for Comparative Economic Studies)
Chairman: ABRAM BERGSON, Harvard University
Papers: GUR OFER, Hebrew University
Industrialization with Low Level Urbanization, The Experiences of Socialist Eastern Europe
GEORGE STALLER, Cornell University
Postwar Industrial Growth in Czechoslovakia
ALAN BROWN, University of Windsor and JOSEPH A. LICARI, Occidental College
Economic Growth, Stability, and Structural Change in Post-War Hungary
Discussants: MICHAEL MONTIAS, Yale University
BELA BALASSA, Johns Hopkins University

10:30 A.M. THE WORLD DYNAMICS OF ECONOMIC GROWTH (Joint Session with the American Statistical Association)
Chairman: HAROLD BARGER, Columbia University
Papers: RICHARD F. KOSOBUD AND WILLIAM D. O'NEILL, University of Illinois at Chicago Circle
Sensitive Growth Models and Robust Conclusions

* Not to appear in the *Papers and Proceedings*

HANS LINNEMANN, Free University of the Netherlands
Perspectives on a New World Economic Model
HERMAN DALY, Louisiana State University
The Economics of the Steady State
WILLIAM D. NORDHAUS, Yale University
Are There Ultimate Limits to Economic Growth?
Discussants: TJALLING C. KOOPMANS, Yale University
LEONARD ROSS, Columbia University Law School

10:30 A.M. NEW DIMENSIONS OF ECONOMIC INEQUALITY
Chairman: ROBERT J. LAMPMAN, University of Wisconsin
Papers: JAMES D. SMITH, Pennsylvania State University
The Concentration of Personal Wealth, 1953–1969
JOSEPH A. PECHMAN AND BENJAMIN OKNER, The Brookings Institution
What Does the Overall Tax System Do to the Distribution of Income?
LEONARD HAUSMAN, Brandeis University
An Appraisal of Income-Conditioning of Benefits as a Redistributive Strategy
Discussants: HOWARD M. WACHTEL, American University
HOWARD TUCKMAN, Florida State University
GUY H. ORCUTT, Yale University

10:30 A.M. INTERNATIONAL TRADE AND FINANCE III: BALANCE OF PAYMENTS AND INTERNATIONAL INVESTMENT* (Joint Session with the American Finance Association and the American Statistical Association)
Chairman: LAWRENCE B. KRAUSE, The Brookings Institution
Papers: LAWRENCE E. KLEIN, University of Pennsylvania
LINK Model Simulations of International Trade: An Evaluation of the Effects of Currency Realignments
PETER B. KENEN, Princeton University
The Balance of Payments and Policy Mix: Simulations of an Empirical Model of the United States
MICHAEL ADLER, Columbia University AND GUY V. G. STEVENS, Federal Reserve Board
Direct Investments of Multinational Firms and U.S. Exports
Discussants: C. FRED BERGSTEN, The Brookings Institution
HELEN JUNZ, Federal Reserve Board

10:30 A.M. ASPECTS OF PROPERTY TAXATION* (Joint Session with the American Real Estate and Urban Economics Association)
Chairman: DONALD PHARES, University of Missouri-St. Louis
Papers: JOHN CALLAHAN AND WILLIAM WILKEN, Georgia State University
Role of Property Tax Reform in School Finance Revisions
THOMAS IRELAND AND WILLIAM MITCHELL, University of Missouri-St. Louis
Circuit-Breaker Legislation: Public Choice Effects on Community Selection
ROY BAHL AND DAVID GREYTAK, Syracuse University
The Property Tax and the Changing Structure of City Employment: Revenue Implications
Discussant: ROBERT SAUNDERS, Kent State University

1:30 P.M. ECONOMIC ANALYSIS OF FEDERAL HEALTH POLICIES* (Joint Session with the Health Economics Research Organization)
Chairman: DONALD E. YETT, University of Southern California
Papers: JOHN E. KOEHLER AND ALBERT P. WILLIAMS, JR., Rand Corporation
Economic Implications of Changes in Funding Mechanisms for Medical Education
PAUL J. FELDSTEIN, University of Michigan
Health Professions Educational Assistance Act As It Applies to Dentistry
STUART H. ALTMAN AND JOSEPH EICHENHOLZ, Cost of Living Council
Impact of the Economic Stabilization Program in Controlling Health Costs
Discussants: REUBEN A. KESSEL, University of Chicago
ALEX R. MAURIZI, U.S. Department of Labor
DAVID DRAKE, American Hospital Association
CHRIS N. THEODORE, American Medical Association

* Not to appear in the *Papers and Proceedings*

1:30 P.M. SUBSIDIES IN INTERNATIONAL TRADE* (Joint Session with the Association for the Study of the Grants Economy)

Chairman: MARTIN PFAFF, Wayne State University and University of Augsburg

Papers: KANTA MARWAH, Carleton University and U.N.C.T.A.D.

Toward a Generalized Model of Export Subsidies

JANOS HORVATH, Butler University

On the Leverage of Subsidies as Policy Tools in East-West Trade Promotion

SIGHART NEHRING, University of Augsburg AND FRANK WEISS, University of Augsburg and University of Pennsylvania

Domestic Transfers and Price Distortions in International Trade

Discussants: BELA BALASSA, Johns Hopkins University

WITOLD TRZECIAKOWSKY, Foreign Trade Research Institute, Warsaw

GERHARD FELS, Weltwirtschaftsinstitut, University of Kiel

1:30 P.M. THE NEGATIVE INCOME TAX EXPERIMENTS: A REPORT OF THE MAJOR FINDINGS* (Joint Session with the Econometric Society)

Chairman: JAMES TOBIN, Yale University

Papers: GLEN CAIN AND HAROLD WATTS, University of Wisconsin

Basic Labor Supply Response Findings from the Urban Experiment (New Jersey-Pennsylvania)

CHARLES METCALF, University of Wisconsin

Low Income Households and the Permanent Income Hypothesis: Implications of the Urban Experiment

D. LEE BAWDEN, University of Wisconsin

Economic Behavior of Households in the Rural Experiment (North Carolina-Iowa)

Discussants: ALBERT REES, Princeton University

ROBERT LEVINE, Rand Corporation

* Not to appear in the *Papers and Proceedings*

ANNOUNCEMENTS

Members with suggestions for the meetings in San Francisco, December 28–30, 1974, are invited to send them to Rendigs Fels, Secretary-Treasurer, American Economic Association, 1313 21st Avenue, South, Nashville, Tennessee 37212. The Secretary-Treasurer will review the suggestions and forward them to the Program Chairman. Papers at sessions sponsored by the American Economic Association are ordinarily invited rather than contributed. To be considered, proposals for this Association must be received by the Secretary-Treasurer no later than February 1, 1974. Economists wishing to give papers on econometrics or economic theory may want to submit them to the Econometric Society, which meets with the Association and annually schedules a number of contributed papers.

As many members of the American Economic Association will be aware, the London School of Economics has recently had the opportunity of purchasing a large adjacent building at modest cost which will greatly increase its library facilities. The library of the School is one of the two or three best collections in the world. It is also a collection that is extensively used by American scholars. The British Government has made a contribution to the effort; Sir John Hicks is contributing his Nobel Prize money; a widespread appeal is being launched among the British public; and, quite properly, American economists are being invited to give—a substantial amount if possible, a small contribution in any case. We would like to join in urging all members of the Association to support this potentially valuable resource of our profession. Contributions are tax deductable if made through the American Friends of LSE, P.O. Box 2033, Princeton, New Jersey 08540. Signed Kenneth J. Arrow and John Kenneth Galbraith.

Nominating Committee of The American Economic Association

In accordance with Section IV, paragraph 2, of the bylaws of the American Economic Association as amended in 1972, President-elect Walter W. Heller has appointed a Nominating Committee for 1974 consisting of James Tobin, Chairman; Allen C. Kelley, George Perry, Roy Radner, Christopher Sims, Alice Vandermeulen, and Charles Z. Wilson. Attention of members is called to the part of the bylaw reading, "In addition to appointees chosen by the President-elect, the Committee shall include any other member of the Association nominated by petition including signatures and addresses of not less than 2 percent of the members of the Association, delivered to the Secretary before December 1. No member of the Association may validly petition for more than one nominee for the Committee. The names of the Committee shall be announced to the membership immediately following its appointment and the membership invited to suggest nominees for the various offices to the Committee."

The Executive Committee of the American Economic Association at its meeting on March 2, 1973 voted to increase dues 5 percent effective January 1, 1974, subject to reconsideration at its meeting December 27, 1973. This action was taken under the provision of the bylaws permitting the Executive Committee to increase dues proportionate to increases in relevant costs and prices.

The Asia Foundation has awarded the American Economic Association a grant to assist graduate students and visiting professors from Asia currently at institutions East of the Mississippi to attend the annual meeting of the Association. The 1973 meeting will be held in New York, New York, December 28–30. Approximately sixteen travel grants up to $150 each will be awarded. Inquiries should be addressed to the American Economic Association, 1313 21st Avenue South, Nashville, Tennessee 37212.

On Thursday, May 16, 1974, the economics department of the City College of New York will sponsor an all-day conference on "The Economic Analysis of Public Decision Making and Public Processes." The object of the conference is to stimulate research that is relevant to current public issues and policies. Persons interested in reading papers at the conference are invited to submit them to Professor Robert D. Leiter or Professor Gerald Sirkin, Department of Economics, City College of the City University of New York, Convent and West 138th Street, New York, N.Y. 10031. Papers delivered at the conference are scheduled for publication in the second annual conference volume of the department.

Omicron Delta Epsilon, the International Honor Society in Economics, invites the submission of entries for the sixth year of the Irving Fisher and Frank W. Taussig Award Competitions. The Fisher Award consists of $1,000 and publication as a monograph, to be published by Princeton University Press, or a journal article, to be published in the *American Economic Review*. The Taussig Award consists of $100 and publication in *The American Economist*.

Entries for the Fisher Award should be submitted to the Departmental Selection Committees by January 1, 1974, and entries for the Taussig Award by May 15, 1974. They will be judged by the International Editorial Board and finalists by the Final Selection Board, consisting of Professors William J. Baumol, Frank H. Hahn, Leonid Hurwicz, Wassily W. Leontief, and Egon Neuberger (editor). For more information, write to Egon Neuberger, Editor, Economic Research Bureau, State University of New York, Stony Brook, New York 11790.

At the invitation of the Government of the Republic of Austria, the Sixth International Conference on Input-Output Techniques will be held in Vienna, from April 22 to 26, 1974, at the Hofburg Conference Centre. The conference is sponsored by the United Nations and Harvard and Brandeis Universities. Conference languages will be English, French, and Russian. Registration must be completed no later than December 31, 1973. For further information, write to Sixth International Conference on Input-Output Techniques, Postfach 91, A-1103 Vienna, Austria.

American scholars scheduled to participate in international meetings abroad may, as in the past, receive a travel grant from the American Council of Learned Societies. Under the new procedures adopted by ACLS effective October 1973, the applicant must obtain a standard application form from the appropriate society and return the completed form to the society by the applicable deadline: February 15 (for meetings scheduled June to September), June 15 (for meetings scheduled October to January), or October 15 (for meetings scheduled February to May). Applicants will be carefully screened by the society. ACLS instructions state: "*Awards are* restricted to humanists and to those social scientists and legal scholars who are *strongly* oriented toward the humanities, who use humanistic methods in their research, and who will be participating in meetings concerned with the humanistic aspects of their disciplines." Inquiries should be addressed to Rendigs Fels, Secretary-Treasurer, American Economic Association, 1313 21st Avenue South, Nashville, Tennessee 37212.

The Afghanistan Studies Association has recently been organized to promote research, cooperation, contact, and mutual support among scholars committed to the study of Afghanistan. The Association is affiliated with the AAS. Its initial objective is to bring interested scholars into effective contact and to stimulate greater participation and support for a largely neglected field. Further information on the Association activities and goals can be obtained from the Executive Committee, Richard S. Newell, Chairman, History Department, University of Northern Iowa, Cedar Falls, Iowa 50613.

Deaths

Emile Despres, Stanford University, Apr. 23, 1973.

Robert M. Lichtenberg, economic analyst, Port Authority of New York and New Jersey, May 28, 1973.

N. Arnold Tolles, professor emeritus, New York School of Industrial and Labor Relations, Cornell University, Apr. 10, 1973.

Paul Wiers, retired Federal Civil Service employee, Mar. 12, 1973.

Retirements

Helen A. Cameron, associate professor of economics, Ohio State University, June 1973.

William J. Fellner, professor of economics, Yale University, June 1973.

Raymond W. Goldsmith, professor of economics, Yale University, June 1973.

Emile Grunberg, department of economics, University of Akron.

A. Stuart Hall, professor, department of economics, University of Nebraska-Lincoln, June 1973.

Thor Hultgren, professor emeritus, department of economics, University of Wisconsin-Milwaukee, May 1973.

Clifford L. James, professor of economics, Ohio State University, June 1973.

Edmund A. Nightingale, professor of economics and transportation, University of Minnesota, June 1972.

Peter F. Palmer, professor emeritus, department of economics, California State University, Long Beach, 1972.

Visiting Foreign Scholars

Christopher I. Higgins, Australian Treasury visiting lecturer, University of Pennsylvania, 1973–74.

James Maycock, economic adviser, Midland Bank, London: visiting scholar, department of banking and finance, University of Georgia, spring quarter 1974.

Promotions

Virote Angkatavanich: professor of economics and finance, Fairleigh Dickinson University, Sept. 1973.

Marcelle V. Arak: special assistant, research and statistics function, Federal Reserve Bank of New York.

Mojmir Bednarik: professor of economics and finance, Fairleigh Dickinson University, Sept. 1973.

Bernard P. Bernsten: senior economist, finance department, Economic Analysis Division, U.S. Postal Service Headquarters, Oct. 1972.

Jack L. Bishop, Jr.: manager, management sciences, Kentucky Fried Chicken Corporation, Oct. 1, 1972.

Joseph S. Chung: professor of economics, Illinois Institute of Technology, Sept. 1973.

Charlel L. Cole: professor of economics, California State University, Long Beach.

Simeon J. Crowther: associate professor, department of economics, California State University, Long Beach.

Eugene A. Diulio: associate professor economics, Fordham University, Sept. 1973.

Evangelos Djimopoulos: associate professor of economics and finance, Fairleigh Dickinson University, Sept. 1973.

Susan B. Foster: chief, Balance of Payments Division, Federal Reserve Bank of New York.

Constantine Glezakos: associate professor, department of economics, California State University, Long Beach.

Herschel I. Grossman: professor of economics, Brown University, July 1, 1973.

W. Clayton Hall: associate professor of economics, Illinois Institute of Technology, Sept. 1972.

Thomas F. Head: assistant professor of economics, George Fox College, Sept. 1973.

John S. Hill: chief, Securities Analysis Division, Federal Reserve Bank of New York.

Fred H. Klopstock: adviser, foreign department, Federal Reserve Bank of New York.

F. Charles Lamphear: professor of economics, University of Nebraska-Lincoln, Sept. 1973.

Craig R. MacPhee: associate professor of economics, University of Nebraska-Lincoln, Sept. 1973.

William D. Nordhaus: professor of economics, Yale University, July 1973.

Donald O. Parsons: associate professor of economics, The Ohio State University.

Robert J. Rohr: associate professor of economics, Brown University, July 1, 1973.

Alan M. Rugman: assistant professor of economics, University of Winnipeg, Sept. 1973.

Harl E. Ryder, Jr.: professor of economics, Brown University, July 1, 1973.

Bernard Saffran: professor, department of economics, Swarthmore College.

V. V. Sharma: associate professor of economics, Ohio State University, Mansfield Regional Campus.

Elbert W. Segelhorst: professor of economics, California State University, Long Beach.

James A. Taylor: associate professor, department of economics, New Mexico State University, Aug. 1973.

Sheila L. Tschinkel: manager, securities department, Federal Reserve Bank of New York.

Charles Waldauer: professor of economics, Widener College, July 1, 1973.

H. David Willey: vice president, Loans and Credits, Federal Reserve Bank of New York.

Administrative Appointments

Roy C. Anderson: associate chairman, department of economics, California State University, Long Beach.

Joseph S. Chung: chairman, faculty of economics, Illinois Institute of Technology, Feb. 1973.

E. Gerald Corrigan: secretary, Federal Reserve Bank of New York.

Ronald Crowley, Queen's University: acting director, research programs, Ministry of State for Urban Affairs, Ottawa, Jan. 1, 1973.

Eugenè D. Eaton, Jr.: chairman, division of business, economics, and public administration, University of Alaska, Anchorage, May 1, 1973.

Alexander Garcia: deputy chairman, department of economics and finance, Fairleigh Dickinson University, Sept. 1972.

Jonathan R. T. Hughes: chairman, department of economics, Northwestern University, fall 1972.

Michael C. Lovell: chairman, department of economics, Wesleyan University, July 1, 1973.

Norman N. Mintz: associate dean, Graduate School of Arts and Sciences, Columbia University, July 1, 1973.

Christos C. Paraskevopoulos: chairman, department of economics, Atkinson College, York University, July 1, 1973.

S. Benjamin Prasad: chairman, department of administrative sciences, Ohio University, Feb. 15, 1973.

Rodrigue Tremblay: chairman, department of economics, Université de Montréal, June 1, 1973.

Walton T. Wilford: chairman, department of economics and finance, Louisiana State University, New Orleans, June 1973.

James A. Zwerneman: head, department of economics, and associate director, Center for Business Services, New Mexico State University, July 1973.

Appointments

Elizabeth J. Arnault, Pomona College: assistant professor of economics, Pitzer College, Associated Colleges at Claremont, Sept. 1, 1973.

Micha Astrachan: economist, Market Statistics Division, Federal Reserve Bank of New York.

Wayne Ayers, Cornell University: assistant professor, department of economics, Eastern Michigan University, Sept. 1973.

Costas Azariadis, Carnegie-Mellon University: visiting assistant professor, department of economics, Brown University, 1973–74.

Martin N. Baily, Massachusetts Institute of Technology: assistant professor of economics, Yale University, July 1973.

Mary Ann Baily, Massachusetts Institute of Technology: lecturer in economics, Yale University, July 1973.

Duran Bell, Brookings Institution: economist, management sciences department, The Rand Corporation, July 1973.

Susan Bresler, Humanics, Inc.: assistant professor of economics, University of Tulsa, Sept. 1973.

Thomas Burton, University of Waterloo: research officer, Ministry of State for Urban Affairs, Ottawa, July 1, 1973.

Cheryl A. Cook, Yale University: associate economist, economics department, The Rand Corporation, Washington, Aug. 1973.

Patrick J. Corcoran: economist, Business Conditions Division, Federal Reserve Bank of New York.

Philippe R. DeVille, Stanford University: assistant professor of economics, Whittemore School of Business and Economics, University of New Hampshire, Sept. 1, 1973.

Robert Eagly, University of North Carolina, Chapel Hill: visiting associate professor, department of economics, State University of New York-Binghamton, Sept. 1, 1973.

Matthew D. Edel, Massachusetts Institute of Technology: associate professor of urban studies, Queens College, City University of New York, Sept. 1973.

Brian J. Fabbri: economist, Money and Finance Division, Federal Reserve Bank of New York.

George Fallis, Princeton University: research officer, Ministry of State for Urban Affairs, Ottawa, Sept. 1, 1973.

Stanley Feldman: economist, Market Statistics Division, Federal Reserve Bank of New York.

David E. R. Gay, Texas A&M University: assistant professor of economics, University of Arkansas, Sept. 1973.

Glenn Gotz, Lockheed Aircraft Corporation: associate economist, management sciences department, The Rand Corporation, Feb. 1973.

Robert J. Gordon: professor of economics, Northwestern University.

Bruce C. N. Greenwald: lecturer in economics, Wesleyan University, July 1, 1973.

Joseph Grundfest, London School of Economics: associate economist, management sciences department, The Rand Corporation, Feb. 1973.

Fred S. Hoffman: senior economist and director of energy research, economics department, The Rand Corporation, Feb. 1973.

James Hosek, University of Chicago: associate economist, economics department, The Rand Corporation, Aug. 1973.

Susan Hosek, Northwestern, University: associate economist, economics department, The Rand Corporation, Aug. 1973.

Richard W. Hurd, Michigan State University: assistant professor of economics, Whittemore School of Business and Economics, University of New Hampshire, Sept. 1, 1973.

Katsuhito Iwai, University of California, Berkeley: assistant professor of economics, Yale University, July 1973.

William R. Johnson: visiting lecturer in economics, Wesleyan University, July 1, 1973.

Frederick R. Kaen, Northeastern University: assistant professor of finance, Whittemore School of Business and Economics, University of New Hampshire, Sept. 1, 1973.

Daniel Kazmer: assistant professor, department of economics, Massachusetts Institute of Technology, Sept. 1973.

Robert E. Klitgaard, Harvard University: associate economist, economics department, The Rand Corporation, Feb. 1973.

Cheng-few Lee, State University of New York at Buffalo: assistant professor banking and finance, College of Business Administration, University of Georgia.

Abba P. Lerner, Queens College: professor of economics, Scarborough College, and department of political economy, University of Toronto, July 1973.

David P. Levine: lecturer in economics, Yale Universight, July 1973.

Edward B. Leviton, Wilkes College: economist, Bureau of Economics, Federal Trade Commission, June 1973.

Michael L. Lichstein, Massachusetts Institute of Technology: assistant professor, department of economics, Ohio State University.

William P. Lloyd: assistant professor of banking and finance, College of Business Administration, University of Georgia.

Howard P. Marvel, University of Chicago: assistant professor, department of economics, Ohio State University.

Adele P. Massell, Division of Policy Research, Office of Economic Opportunity: associate economist, economics department, The Rand Corporation, Feb. 1973.

George Miaoulis, Jr., New York University: assistant professor of business administration, Whittemore School of Business and Economics, University of New Hampshire, Sept. 1, 1973.

Marnie W. Mueller: assistant professor of economics, Wesleyan University, July 1, 1973.

Tridib K. Mukherjee, National Institute of Bank Management, Bombay: associate professor of economics, Appalachian State University, Sept. 1, 1973.

F. Patricia Munch, University of Chicago: associate economist, economics department, The Rand Corporation, Aug. 1973.

Gary Nelson, Institute for Defense Analyses: economist, management sciences department, The Rand Corporation, Oct. 1972.

Phillip Nelson: associate professor, department of economics, State University of New York, Binghamton, Sept. 1, 1973.

Marc Nerlove: visiting professor of economics, Northwestern University.

James E. Nickum: lecturer, economics department, California State University, Long Beach, Sept. 1973.

Anton S. Nissen: manager, statistics department, Federal Reserve Bank of New York.

William H. Oakland, Johns Hopkins University: professor, department of economics, The Ohio State University.

John Parknay: economics department, Wells Fargo Bank, N.A., San Francisco, Apr. 1973.

Harold C. Passer, U.S. Department of Commerce: assistant treasurer, Eastman Kodak Company, Rochester, New York, Feb. 22, 1973.

Rolando F. Pelaez: assistant professor, department of economics, New Mexico State University.

William M. Penn, Jr., Ford Motor Company: assistant professor, department of economics, Loyola College, Sept. 1, 1973.

Walter W. Perlick: assistant professor of finance, Colorado State University, Sept. 1973.

Harvey J. Reed: lecturer, department of economics, California State University, Long Beach, Sept. 1973.

Mark R. Rosenzweig, Columbia University: lecturer in economics, Yale University, July 1973.

Ivalee M. Skov: lecturer, department of economics, California State University, Long Beach, Sept. 1973.

V. Kerry Smith, Resources for the Future, Inc.: associate professor, department of economics, State University of New York, Binghamton, Sept. 1, 1973.

Hong K. Sohn: senior lecturer, department of economics, James Cook University, Australia, May 1, 1973.

Hugo F. Sonnenschein: professor of economics, Northwestern University.

John P. Stein, University of Chicago: associate economist, economics department, The Rand Corporation, Mar. 1973.

James M. Suarez: economist, Business Conditions Division, Federal Reserve Bank of New York.

Lorie Tarshis, Stanford University: chairman, Division of Social Sciences, Scarborough College, and professor, department of political economy, University of Toronto, July 1973.

Velma Thompson, University of California, Los Angeles: economist, management sciences department, The Rand Corporation, Feb. 1973.

Khairy, Tourk, University of California, Berkeley: assistant professor of economics, Illinois Institute of Technology, Sept. 1972.

Martin Ulrich, Simon Fraser University: research officer, Ministry of State for Urban Affairs, Ottawa, Apr. 1972.

Finis Welch, City University of New York: profes-

scr of economics, University of California, Los Angeles, Sept. 1973, and senior economist, economics department, The Rand Corporation, Sept. 1973.

John H. Wile, Renesselaer Polytechnic Institute: assistant professor of economics, State University of New York, Stony Brook, Sept. 1, 1973.

J. Holton Wilson, Mount Union College: assistant professor of economics, University of Tulsa, Sept. 1973.

Wynnelle Wilson, Florida State University: visiting assistant professor, department of economics, State University of New York, Binghamton, Sept. 1, 1973.

Daniel Wisecarver, U.S. Department of the Treasury: assistant professor, department of economics, Ohio State University.

Lorene Yap, The Urban Institute: assistant professor, department of economics, State University of New York, Binghamton, Sept. 1, 1973.

Joseph A. Ziegler, Clemson University: assistant professor of economics, University of Arkansas, Sept. 1973.

Leaves for Special Appointments

Trent J. Bertrand, Johns Hopkins University: visiting professor, faculty of economics, Thammasat University, Bangkok, 1973–74.

Benjamin Chinitz, Brown University: visiting professor of economic policy, State University of New York at Binghamton, fall semester 1973.

Avery B. Cohan, University of North Carolina, Chapel Hill: visiting professor of finance, department of banking and finance, University of Georgia, 1973–74.

S. Enke, General Electric: director, Foreign Assistance and Development Task Force, Washington, Mar. 1, 1973.

Michael E. Farbman, University of Glasgow: research officer, World Employment Programme, International Labour Office, Geneva, May 1973–Feb. 1974.

Stanley Lebergott, Wesleyan University: Institute for Advanced Study, Princeton, July 1, 1973.

Seiji Naya, University of Hawaii: visiting profe faculty of economics, Thammasat University, B kok, 1972–74.

Robert J. Rohr, Brown University: visiting fessor, department of economics, Dartmouth Col 1973–74.

Harl E. Ryder, Jr., Brown University: visiting fessor, department of economics, Stanford Univer 1973–74.

Resignations

A. Gordon Ball, Iowa State University: Unive of Guelph.

Michael R. Darby, Ohio State University: Unive of California, Los Angeles, Feb. 1973.

William H. Foeller, University of Akron: N eastern State College, Sept. 1973.

Marsha G. Goldfarb, Yale University, June 197

Robert S. Goldfarb, Yale University: George W ington University. June 1973.

John J. Hooker, University of Texas-El Paso: Worth National Corporation, Apr. 16, 1973.

F. Thomas Juster, National Bureau of Econ Research: University of Michigan, May 1, 1973.

Stephen E. McGanghey, Iowa State Univer International Development Bank.

John J. McGowan, Yale University: Charles F Associates, Inc., Cambridge, Mass., June 1973.

Prithvi R. Mathur, University of Akron.

John Peter Mattila, Ohio State University: I State University, Sept. 1973.

Richard D. Porter, Ohio State University: Fe Reserve, June 1973.

James W. Robinson, University of Akron: Vir Polytechnic Institute and State University, Sept. 1

James T. Turner, University of Akron: Packa Corporation of America, June 1973.

NOTE TO DEPARTMENTAL SECRETARIES AND EXECUTIVE OFFICERS

When sending information to the *Review* for inclusion in the Notes Section, please use the following style:

A. Please use the following categories:

1—Deaths
2—Retirements
3—Foreign Scholars (visiting the USA or Canada)
4—Promotions
5—Administrative Appointments
6—New Appointments
7—Leaves for Special Appointments (NOT Sabbati
8—Resignations
9—Miscellaneous

B. Please give the name of the individual (SMITH, John W.), his present place of employment or enrollment new title (if any), his next place of employment (if known or if changed), and the date at which the change will o

C. Type each item on a separate 3x5 card, and please do not send public relations releases.

D. The closing dates for each issue are as follows: *March*, November 1; *June*, February 1; *September*, Ma *December*, August 1.

This announcement supersedes and replaces a letter which was sent annually from the managing editor's offic items and information should be sent to the assistant editor, *American Economic Review*, Box Q, Brown Univer Providence, Rhode Island 02912.

The American Economic Review

ARTICLES

DECEMBER 1973

THE AMERICAN ECONOMIC ASSOCIATION

Founded in 1885

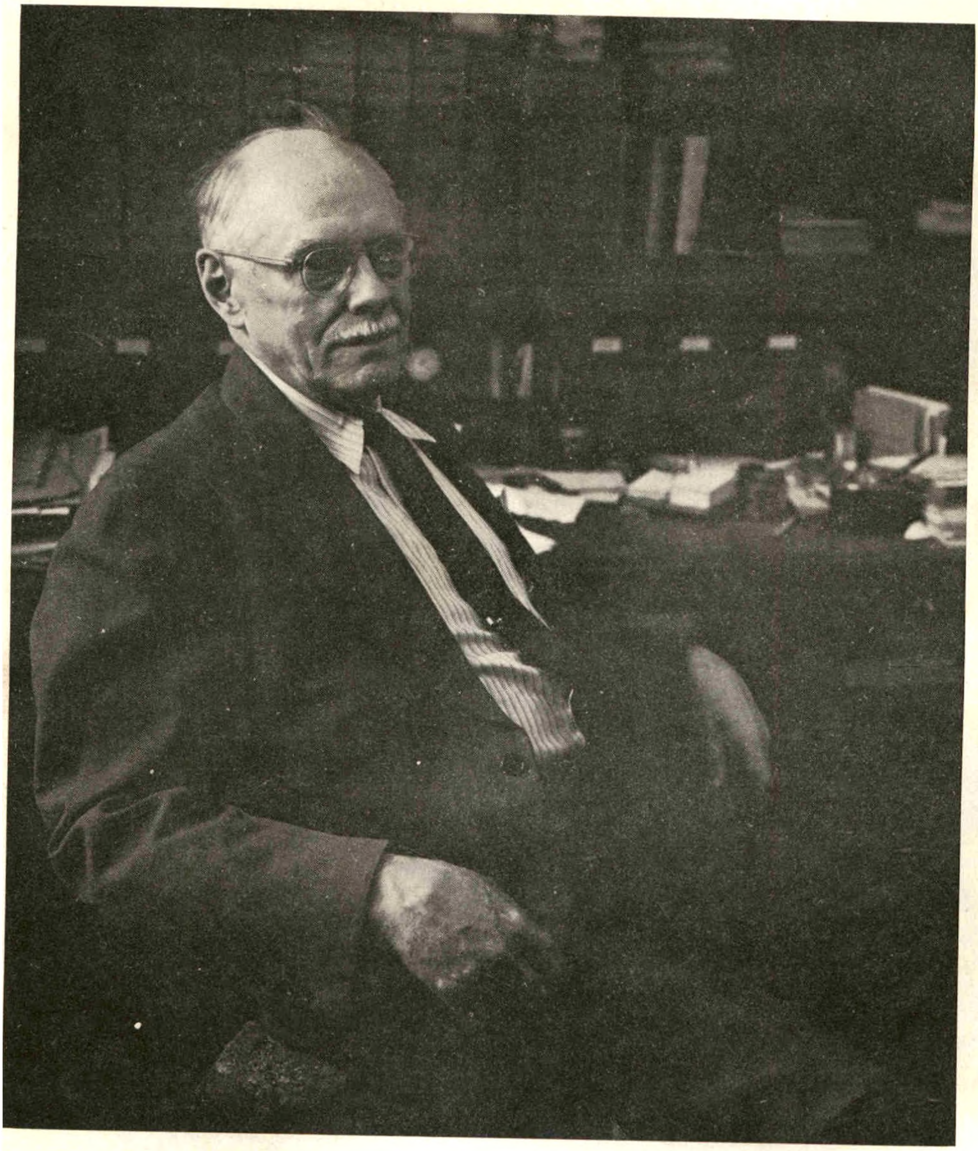
• Published at George Banta Co., Inc., Menasha, Wisconsin.

• THE AMERICAN ECONOMIC REVIEW, including four quarterly numbers, the *Proceedings* of the annual meetings, and *Directory* and Supplements, is published by the American Economic Association and is sent to all members five times a year, in March, May, June, September, and December.

• Membership dues of the Association are $21.00 a year, which includes a year's subscription to both the *American Economic Review* and the *Journal of Economic Literature*. Junior membership at $10.50 per year is available to registered students for a maximum of three years. Subscriptions (libraries, institutions, or firms) are $31.50 a year, and only subscriptions to both publications will be accepted. Single copies of the *Review* and *Journal* are $4.50 each plus a $.50 per order service charge. Orders should be sent to the Secretary's office, Nashville, Tennessee.

• Correspondence relating to the *Papers and Proceedings*, the *Directory*, advertising, permission to quote, business matters, subscriptions, membership and changes of address should be sent to the Secretary, Rendigs Fels, 1313 21st Avenue South, Nashville, Tennessee 37212. Change of address must reach the Secretary by the 1st of the month prior to the month of publication. The Association's publications are mailed second class and are not forwarded by the Post Office.

• Second-class postage paid at Nashville, Tennessee and at additional mailing offices. Printed in U.S.A.



Frank Knight at 70

Frank Knight as Teacher

By DON PATINKIN*

I shall write of Frank Knight[1] as a teacher—and this is really the viewpoint from which I can best describe him. For I do not feel that I have the necessary expertise in the theory of uncertainty and in capital theory—and certainly not in social philosophy—to provide a critical evaluation of Knight's fundamental contribution to these fields. Nor do I have any basis to write of him as a colleague. On the other hand, my most vivid memories of Knight even today are from the days when he was one of my teachers at the University of Chicago (1941–47). Indeed, I saw him afterwards only a few times. Furthermore, my own work in economics has largely been in fields that were not his major concern, and so I read—or reread—little of his work after my student days.

Thus the picture of Knight that I shall draw here is primarily the one that I remember seeing as a student—through the window on the life of his teacher that is open to a student. It is thus a picture of only one aspect of Frank Knight's life—though obviously an important one. And I am fully aware that—like any picture based on personal impressions and memories—it is one that also reveals something about the viewer.

What is the mark of a great teacher? It is first and foremost the qualities he con-

* Professor of economics, The Hebrew University of Jerusalem. Appropriately enough, the first draft of this memoir was written while serving during the fall of 1972 as Ford Foundation Visiting Research Professor of Economics at the University of Chicago. And though the pages which follow are by their very nature a personal document, I have nevertheless greatly benefited from—and am grateful for—the opportunity afforded by this visit of discussing them with D. Gale Johnson, T. W. Schultz, George Stigler, Roger Weiss, and other people at Chicago who knew Frank Knight well. I am also indebted for helpful comments on earlier drafts to Yoram Ben-Porath, Martin Bronfenbrenner, Clarence Efroymson, Richard Freeman, William Ginzberg, Giora Hanoch, Ephraim Kleiman, Leonardo Leiderman, Akiva Offenbacher, David Marawetz, Jerome Stein, Lester Telser, Menahem Trajtenberg, and Menahem Yaari. In view of the personal nature of this memoir, it is really unnecessary to say that none of these individuals bear responsibility for the description that follows.

I would also like to express my appreciation to George Stigler for making Knight's unpublished papers (which are in his custody) available to me, and to Glen Gilchrist (who has catalogued these papers under Stigler's direction) for guiding me through them and xeroxing requested materials. I wish finally to thank Mark Wang at the University of Chicago and Akiva Offenbacher at the Hebrew University of Jerusalem for their valuable technical assistance, and Kathryn Bates and, especially, Vera Jacobs for patiently and accurately typing this memoir through its various drafts. I am also grateful to the University of Chicago, the Central Research Fund of the Hebrew University, and the Israel Academy of Sciences and Humanities for research grants that made this and other assistance possible.

With regard to articles listed in the reference section, page numbers shown in the text refer to the reprinted version, and not to the original.

[1] Frank Hyneman Knight: born, McLean County, Illinois, November 7, 1885; died, Chicago, Illinois, April 15, 1972. The bare facts of Knight's academic career as given by the 1969 *Handbook of The American Economic Association* are as follows: Ph.B., Milligan College, Tennessee, 1911; B.S., M.A., University of Tennessee, 1913; Ph.D., Cornell University, 1916; instructor of economics, Cornell University, 1916–17, University of Chicago, 1917–19; associate professor and later professor of economics, State University of Iowa, 1919–27; professor of economics, University of Chicago, 1927–45; distinguished service professor of social sciences and philosophy, 1945–52; emeritus after 1952.

A bibliography of Knight's writings for the period 1915–35 appears at the beginning of his *Ethics of Competition* (1935, pp. 11–18). For other details of Knight's life, as well as a general description of his work, see James Buchanan; see also Joseph Dorfman, pp. 467–79, 767–70, *et passim*). On Knight's days as a graduate student at Cornell—and his shift then from studying philosophy to economics—see the description of his renowned teacher there, Alvin Johnson (1960, pp. 227–28, 231). (I am indebted to Edward Shils for this reference.) See also the biographical sketch of Knight that appears elsewhere in this issue.

veys by his very presence in the classroom: his personal integrity and his intellectual curiosity and stimulation; his humility and his breadth of interests.

And it is, secondly, the insights and understanding—the new ways of looking at things—that he transmits to his students. Frequently these insights are original to him; but even when they are not, they reflect his judgment as to what is important in the existing body of knowledge, and hence worthy of emphasis. And it is the mark of a great teacher that the insights he thus passes on to his students do indeed remain important: that they continue to guide their thinking many years later; and that these students in turn consider the insights so important as to wish to pass them on to their own students as well.

In all of these ways Frank Knight was a great teacher.

I am afraid that I did not think so in my first contact with him when I tried sitting in on his course on "Price and Distribution Theory" (Economics 301), the first course in economic theory required of all graduate students. I found myself then quite confused by the middle-aged (he was then in his late 50's), medium-height, plumpish and moustached man who stood at the side of the large elliptical wooden table in one of the seminar rooms on the first floor of the Social Sciences Research Building—leaning on the back of a chair, occasionally puffing on a corn-cob pipe—and rambling on in a high-pitched voice and in a disjointed manner on mysterious issues that certainly cast no light on the newly revealed truth which was then being enthusiastically explicated everywhere—in words, in graphs, as well as in mathematical formulas—of "marginal revenue = marginal cost." And after a few such bewildering experiences, I gave up in despair.

But not for long. The following year I registered for the course, and this time succeeded in taking fairly coherent classnotes. But I must admit that my greatest pleasure and benefit from the course came when, toward the end of my graduate studies, I sat through it once again.

Many factors lie behind this long road to understanding and appreciation. First of all, Knight gave little emphasis in his teaching to those things to which the beginning graduate student is normally attracted—namely the technical aspects of the discipline, and the newer the better. In part this was due to the fact that Knight was just not interested in these aspects—and in part because he took this knowledge for granted and wanted to get at the more fundamental issues that lay behind the assumptions and implications of the analysis. And this was not simply a reflection of the period and of the generation—for my classnotes of Jacob Viner's contemporaneous version of the Economics 301 course show that Viner (in addition to being concerned with the broader issues of analysis and scholarship) gave considerably more emphasis to the technical aspects of the analysis than did Knight. Thus, for example—as might be expected from his famous article on "Cost Curves and Supply Curves" (1931)—Viner provided a fairly detailed presentation of the properties of these curves under the assumptions of imperfect as well as perfect competition, though (as I shall note later) Knight too discussed the theory of imperfect competition.

But this was not the only cause of the beginning graduate student's difficulties with Knight's course. For, quite frankly, Knight was not a good teacher in the sense of systematically introducing and developing a subject. Nor did he make a pedagogic effort to motivate the student to understand the subject in question by explicitly relating it to the general framework of economic analysis. All this, Knight took for granted as being known to the student—

and devoted himself instead to forcing the student to rethink the basic issues of economic theory as he saw them.

Another difficulty with Knight's lectures was that at crucial points they frequently (and unawares to the student) turned into brief and cryptic summaries of views that Knight had developed at length in various of his writings—to which for the most part Knight did not explicitly refer. Thus a brief reference to the invalidity of "productivity ethics" or to "commutative justice *versus* distributive justice" could hardly convey the depth and meaning of Knight's famous essay on "The Ethics of Competition." Nor could passing references to "real-cost versus alternative-cost theories of value" mean much to a student who was not aware of Knight's many writings on this question.

For this reason I remember characterizing Knight's lectures at that time as being like a general-equilibrium system—which could not be solved until the student was familiar with the set of Knight's articles that specified the relationships between the various parts of the lectures, and thus converted their unknowns into knowns. And I still remember my feeling of satisfaction when I began to be able to see the system of equations as a whole—and things began to fall into place.

There is one further bit of background information that I feel is relevant to my picture of Knight. The teachers of economic theory during my student days at Chicago included not only Frank Knight and Jacob Viner, but also Oskar Lange—who was the antithesis of Knight, and not only on political grounds. Where Knight devoted much attention to probing (in his rambling and often obscure manner) into the meaning of the basic definitions and assumptions of the analysis ("perfect competition," "perfect foresight," "wants," "costs," "capital," "equality," and the like), Lange (in his contrastingly clear and systematic manner) was primarily concerned with drawing the logical implications of these assumptions. So where Knight taught economic theory in a loose, "literary," philosophical fashion—and was antipathetic to mathematical economics[2]—Lange was formal, rigorous, complete, and frequently made use of mathematical tools (not to mention his invaluable course in Mathematical Economics). And where Knight was basically not sympathetic to the new developments in economic theory (read: Keynes and Hicks)—and even, I would say, instinctively critical of them—Lange was an early convert as well as an efficient expositor and refiner.[3] Thus Knight and Lange complemented each other in a most wondrous way—thereby increasing the productivity of each of them in the teaching process. From the implicit dialogue that thus took place between these two teachers, we students were the direct beneficiaries.[4]

[2] "The mathematical economists have commonly been mathematicians first and economists afterward, disposed to oversimplify the data and underestimate the divergence between their premises and the facts of life. In consequence they have not been successful in getting their presentation into such a form that it could be understood, and its relation to real problems recognized, by practical economists" (Knight 1923, p. 49).

[3] See Lange's 1938 model of the *General Theory* in his "The Rate of Interest and the Optimum Propensity to Consume" and his use and elaboration of the analytical tools of *Value and Capital* in his *Price Flexibility and Employment* (1945).

[4] Unfortunately, neither side to this dialogue—and even more unfortunately, practically nobody at Chicago at that time—represented the empirical approach to economics in the sense of the statistical analysis of quantitative data. From the death of Henry Schultz in 1938 and until the advent of the Cowles Commission in the mid-1940's (whose major concern at that time was not only the methodology of econometric models, but also the actual construction of such models for the *U.S.* economy), this component of a good education in economics was almost entirely absent from Chicago. It should, however, be noted that empirical economics had not achieved the recognized position it now has in the profession, and was accordingly absent from many (if not most) of the other leading academic centers as well.

I. Knight the Theoretical Economist

What were the main features of this world of Knight's? It was first and foremost ("If you don't learn anything else from me, then learn this . . .") the notion of social organization—and the economic system as a means of fulfilling certain basic functions of any such organization: the system whose exact structure varies from society to society, but which for every one of them must answer the basic questions about what goods and services to produce, how to produce them, how to distribute them among the members of the society, and how to provide for the future of the society. The fruitfulness of this view of economics is well-attested by the fact that it has been adopted by such widely used elementary textbooks as those of Samuelson, Lipsey, and others, as to have become by now a commonplace.[5]

Another basic element of Knight's teaching was the view that the study of man—including economic man—could not proceed within the same deterministic framework as the study of nature. For the behavior of man must also express his freedom of will—to which there is no counterpart in the physical world. Hence "fully determined" or "perfectly rational" human behavior is no behavior at all: for such behavior is entirely mechanistic, leaving no place for free choice on the part of the individual. ("Does the apple falling from the tree act rationally? The question is meaningless—since the apple is subject to mechanical forces.") Indeed, the proper study of human behavior must take into account the possibility of error and the existence of exploratory and game-playing aspects of such behavior. Correspondingly, despite his occasional use of analogies from physics, Knight emphasized that we "cannot transfer the laws of the physical sciences to where human beings are concerned" (Classnotes: Econ. 305, 1945).

In the opening lecture of his theory course in 1942, Knight (according to my classnotes) set out his general approach to the analysis of the economic system:

> The main theme of traditional economics is that the investigation of theoretical problems can be settled by argument, without using inductive argument. Discuss things of common knowledge. Most questions solve themselves if correctly stated. No special technique necessary. The current tendency in the literature is a reaction against this view. Want to use inductive thinking—getting at facts. Also use mathematical techniques—statistics. Point of view of this course is not inductive or deductive in the sense of being beyond reach of any educated person. Simple principles of mathematics to be used. [Classnotes, Econ. 301, 1942]

Our basic readings for the course were from Marshall's *Principles* and Davenport's *Economics of Enterprise*.[6] In the

[5] Knight discussed these basic functions in his 1933 booklet on *The Economic Organization*. This reproduces the four chapters that Knight first published in 1932 in the book of readings that was prepared specially for the famous social science survey course developed during the 1930's at the College of the University of Chicago. In the bibliography of Knight's work that appears in the *Ethics of Competition* (1935c, p. 15) these chapters are described as "an abstract of material mimeographed for private circulation at the University of Iowa during the years 1922–25." (This is apparently the material that has been found among Knight's papers and catalogued under the title "*Economics*, 1920's." In any event, this material does contain an approach to economic analysis in terms of the functions of an economy.) For further details on the dating of this material, see fn. 1 in my paper elsewhere in this *Review*. An early version of the discussion of the functions of an economy is to be found in Knight's doctoral thesis, as revised and published in his classic *Risk, Uncertainty and Profit* (1921a, pp. 54 ff). In connection with Samuelson's approach to economics in terms of these functions, see his reference to Knight in the second edition of *Economics: An Introductory Analysis* (1951, p. 14, fn. 1). See also fn. 8 below.

[6] The other reading materials to which we were referred (according to my 1942 classnotes from Knight's 301 course) were as follows: E. H. Chamberlin, *Monopolistic Competition* (1942) (mentioned, but not clear if assigned as reading); R. F. Harrod, "Doctrines of Imperfect Competition" (1934); J. R. Hicks, *Value and Capital*, ch. 1 (1939); F. H. Knight, *The Economic Or-*

opening lecture just mentioned, Knight referred in the following words to these writers and to what he planned to do in the course:

> Marshall was in many ways an eclectic combining supply and demand sides. He was a conservative and a defender of *laissez-faire*. Davenport was one of the most important and original thinkers.[7] He was a radical in the negative sense. . . . Davenport was a forerunner of Keynes. In certain paragraphs one can't tell whether reading Keynes or not.
>
> Knight is for reading a couple of books and getting conclusions of subject rather than reading literature—what people have said (as Viner does). Knight interested in social applications of economic approach; not interested in associating points of view with writers, or in puzzle solving. The big job of economics is to divest people of prejudices—to have them see the questions as they are. Right now ordinary lay opinion should be forgotten as much as possible [by the students, presumably]. [Classnotes: Econ. 301, 1942]

The primary contrast that Knight drew in his lectures—and in much of his writings over the years—was that between two alternative systems for fulfilling the aforementioned functions: the competitive (capitalist) system and socialism. Most of his classroom analysis (as well as of his writings) was devoted to the competitive market system. Here Knight gave an overview of this system in terms of what he called "the familiar figure of the 'wheel of wealth'," (1933a, pp. 60–61, here reproduced in Figure 1) which has become even more familiar to a modern generation of economists through its use by Paul Samuelson in his introductory textbook.[8] Within the framework of this

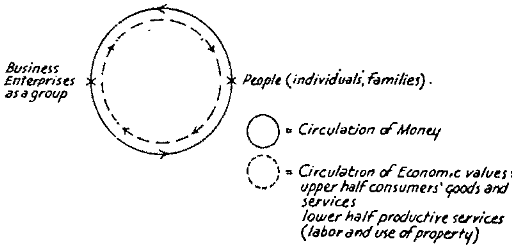


FIGURE 1

ganization (1933a); A. P. Lerner, "Statics and Dynamics in Socialist Economics" (1937). In my 1945 classnotes, there is also a reference to Knight's "The Business Cycle, Interest and Money: A Methodological Approach" (1941).

Other of Knight's writings that we read in connection with this course included: *Ethics of Competition* (1935d), various essays; "The Quantity of Capital and the Rate of Interest" (1936); "Diminishing Returns from Investment" (1944a); and "Realism and Relevance in the Theory of Demand" (1944b). Interestingly enough, there is only one mention in my notes (in the context of a discussion of the nature of profits) of Knight's classic, *Risk, Uncertainty, and Profit*. Nor do I recall our reading this book as students. On the other hand, it is clear in retrospect that many of Knight's classroom discussions stemmed from it.

[7] The deep impact—both intellectual and (I would conjecture) personal—that Davenport had on Knight is most evident from the biographical sketch of him that Knight wrote for the *Encyclopedia of the Social Sciences* (1931). Knight's first contact with Davenport (which—as the foregoing sketch makes clear—also included participation in the latter's classes) apparently took place at Cornell during 1916–17, which was the year that Davenport came to Cornell (Spiegel, 1968, p. 16), and also the year that Knight stayed on there as an instructor after receiving his Ph.D. (see fn. 1, above).

I might also note that the index to Knight's famous "work of his youth," *Risk, Uncertainty, and Profit*, contains more references to Davenport than to any other writer besides J. B. Clark.

[8] Knight's first use of such a diagram can be dated back to 1926 at the latest (see fn. 5 above). Insofar as Samuelson is concerned, see the circular-flow diagrams that appear in all editions of his *Economics: An Introductory Analysis* (e.g., 1948, p. 226; 1970, pp. 42, 170, and 607). I might note that Samuelson did his undergraduate work at the University of Chicago in the early 1930's—and that in reply to my query as to the possible origin of his circular-flow diagram in Knight's teachings, Samuelson has indeed confirmed that "I [Samuelson] was influenced by Knight's elementary (for him) explanation of how the economic system works. In the Social Science Survey Courses I or II, 1932–3 [see fn. 5 above], I am sure I was assigned it to read. And in early editions of my book, I cited Knight for How, What, and For Whom" (personal letter to author dated April 10, 1973, and cited with Samuelson's kind permission). (For his reminiscences of his days at Chicago, see Samuelson (1972a, pp. 158–61; 1972b, p. 5).)

The question of the origins of Knight's diagram itself is an intriguing one that I have discussed at length in a

diagram, Knight then proceeded to explain how the price system (and more specifically, the two sets of prices—those of final goods and services, and those of productive services) "control[s] the process of production and distribution under free enterprise" (1933a, p. 62).

The distinctive elements of Knight's exposition began with his criticism of the term "competition" as a description of the system. This term, emphasized Knight, had a connotation of *personal* competition: whereas the system of perfect competition was an impersonal one. With whom, asked Knight, was the farmer—producing his agricultural output in a perfect market—competing? Thus the use of the term was misleading in failing to bring out that so-called system of *competition* was a system that actually brought about the *cooperation* (though again, in an impersonal manner) of economic units in carrying out the productive processes of the economy. The basic characteristic of this system—and once again Knight prefaced his discussion here (as in many other contexts!) with the injunction "If you don't learn anything else from me, then learn this . . . "—was that the exchanges which take place under this system are to the *mutual advantage* of both parties to the exchange. Indeed, this—according to Knight—was the major innovation of Adam Smith, though one that was frequently not understood even to this day. And he would frequently refer to the levying of tariffs as a flagrant example of the continued failure to understand this fundamental of economic analysis.

In analyzing the productive process, Knight emphasized that it was one in which (like in the usual production function of today) output was essentially simultaneous with input—and he thus criticized the classical model of (agricultural) production as well as the Austrian production-period model of capital theory, both of which presented output as following inputs with a time lag. Knight also emphasized that essentially what was produced (and, even more so, consumed) was not goods, but services—though (he admitted) the distinction tended to vanish in the case of perishable goods which were immediately destroyed in the process of consumption.

In his analysis of production, Knight devoted much attention to an explanation of the proper meaning of the law of diminishing returns. In this context he made the distinction—now to be found in many price-theory textbooks—between the three regions of the total output curve under the assumption of constant returns to scale (see Figure 2): region I, in which the average product of the variable factor, A, is rising; region II, in which both marginal and average product of A are declining; and region III, in which the marginal product of A is negative. Knight went on to emphasize the symmetry between regions I and III that expresses itself in the fact that just as the marginal product of the variable factor A is negative in region III, so must the marginal product of the fixed factor B be negative in region I. Hence, said Knight, in the real world we are always in region II (see 1921a, pp. 97–102).[9]

Within the foregoing framework Knight forcefully expounded an "alternative" or "opportunity cost" theory of value that he contrasted with the classical "real" or

note that appears elsewhere in this *Review*. This also describes the different uses that Samuelson has made of the circular-flow diagram in the various editions of his textbook.

[9] For examples of textbook discussions of Knight's "three regions," see Stigler (1952, p. 115), Ferguson (1969, p. 130), and Grunfeld, Liviatan, and Patinkin (1963, pp. 15–19). Some price-theory textbooks today make use of the assumption of "free disposal" and thus draw the production curve as *Obce* instead of *Oabcd*, or else simply draw only the relevant segment of the curve, *bc*.

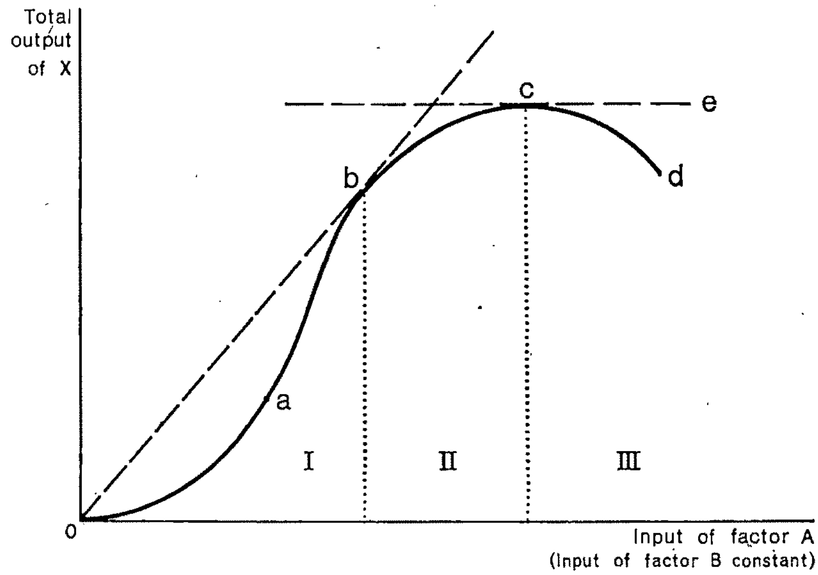


FIGURE 2

"subjective" or "pain cost" theory of value—to the clear disadvantage of the latter. I mentioned before that the beginning student frequently did not understand the issues that Knight was discussing in the classroom until he had found his way to those of Knight's writings that dealt with these issues (and in this case Knight's 1935 article on "The Ricardian Theory of Production and Distribution" was particularly relevant). I might now add that in this case the student's understanding was also advanced in the course of his studies as he became aware (through exposure to Viner's version of the Economics 301 course) that this was not a dead issue, but part of a running debate that Knight was carrying on with Viner. And I must admit that Knight's presentation of the issue in terms of absolute right and wrong (as he was wont to do) was less helpful than Viner's resolution of the debate in terms of including in the individual's calculation of alternatives his subjective tastes for the pains and pleasures he derived from the various labor services he could sell in the market.

In any event, I still find that it deepens one's understanding to follow Knight in emphasizing that Adam Smith's deer was worth two beaver not because of the fact per se that a unit of labor was involved in the catching of each, but because this fact implied that the cost of catching a deer were the two beaver that could alternatively have been caught. Hence (in Knight's terms) the market price for a deer had to be two beaver: for otherwise there would be a discrepancy between what the individual had to forego in order to obtain a (say) deer "directly" (viz., by hunting it) and "indirectly" (viz., by hunting beaver, and exchanging them for the deer), so that market forces of excess supply (or demand) for deer would automatically reestablish the price of two beaver to one deer, should it deviate from this level.

A basic part of Knight's theory course was devoted to the theory of demand, and to the Marshallian demand curve in particular. In this context Knight was less receptive than he might have been to indifference-curve analysis, and there may

indeed be hints of a somewhat more positive attitude to this analysis in his Preface to the 1948 reissue of *Risk, Uncertainty, and Profit* (p. XLVIII). After all, indifference-curve analysis did not challenge anything basic in Knight's general approach, so that I feel that his opposition to it was at least in part a reflection of his reluctance at that time to change traditional modes of thought.

Still, Knight's detailed critique of Hicks' exposition of indifference-curve analysis did provide the student with a deeper understanding of its assumptions. It also enabled Knight to draw a sharp and instructive distinction between the demand curve that Hicks generated by changing the price of the good in question while keeping all other prices and money income constant—so that real income changed along the demand curve; and the Marshallian demand curve, in the generation of which (according to Knight) other prices were changed so as to keep real income constant (1944b, 1946). This interpretation of the Marshallian demand curve was also stressed by Viner in his version of the Economics 301 course (Classnotes, 1944), and so can well be said to be part of a Chicago tradition.[10]

As already noted, Knight discussed (and referred us to the relevant readings on) the theory of monopolistic competition—indeed, to an extent greater than one might infer from some of the things that have been written on the attitude of the Chicago school to this theory.[11] It is, however, true that Knight was critical of Chamberlin's concepts of imperfect competition, and denied that the possibility that an "... enterprise can have a diminishing [i.e., negatively sloped] demand curve but still no monopoly profits ... corresponded to fact" (Classnotes, Econ. 301, 1942).

The most distinctive part of Knight's theory course had to do with his theory of capital—and one cannot but be impressed by the modern flavor of what he taught here. First of all, Knight railed against the traditional classical "trinity of factors of production"—land, labor, and capital. There was little in the productive process that Knight was willing to identify with the "original and indestructible forces of nature" that the classical economists had defined as "land." Agricultural land, too—emphasized Knight—had to be developed and maintained just like any other capital good. Similarly, the productive process of labor reflected primarily not "native qualities," but the artificial qualities which constituted human culture. Thus there is no basic distinction between most of the income received by labor and that received by capital. In the terminology of the more recent "Chicago school" associated with the names of T. W. Schultz, Gary Becker, and others (and I suspect that Knight had some direct or indirect influence on their thinking, too), Knight viewed most of labor income as returns to "human capital" and stressed the role of the family (both genetically and socially) in endowing its children with this capital. Thus Knight was not very far from classifying all of the factors of production under the one general title "capital."

In his more formal treatment of capital theory, Knight—like J. B. Clark before him (1899)—emphasized the basic distinction between specific capital goods in the concrete—which are periodically worn out and reproduced, and the stock of capital in the abstract—which is permanent, and which with the exception of a few historical periods (for example, periods of prolonged warfare) has been continuously

[10] See the similar interpretations—based on texts from Marshall's *Principles*—in Friedman (1949) and Patinkin (1963).

[11] See Archibald (1961) and the references to Stigler and Friedman there cited, as well as the subsequent exchange (1963) between all three of these economists. See also Miller (1962, p. 67).

growing over the course of human history. (See Knight, 1936, p. 434.) And a capital good has value only by virtue of its being the source of the stream of services—and hence of income—that flows from it.

The economic explanation of the rate of interest is to be found in the productivity of capital. Time preference is not a cause of interest, but the result of it; because interest can be earned by deferring present consumption to the future and investing the savings in the meantime, the individual can not be in equilibrium unless a dollar's worth of present consumption exceeded the value to him of a dollar's worth of future consumption by precisely the interest that could alternatively be earned by abstaining now from consumption. I should also note that Knight's emphasis on the fact that capital in the abstract—or wealth—was continuously growing over time led him to reject the Fisherine notion that individuals save in the present only in order to finance increased consumption in the future.

And since the rate of interest is determined by the productivity of capital, it is (as Lerner (1953), Haavelmo, and others were later to emphasize) the total *stock* of capital that is relevant to this analysis—and not the *flow* of current investment, which by its very nature is small relative to the total stock of existing capital, and hence can have little effect on its marginal productivity. (See Knight 1932a, pp. 262–63; 1936, p. 622.) (I might note that this concern with the time dimensions of the various economic variables was a constant characteristic of Knight's analysis—emphasized, for example, also in his treatment of the different "runs" of Marshallian cost and supply curves (Knight, 1921b, pp. 186–216).)

On the basis of this approach to capital theory, and with the aid of his historical perspective, Knight continued to insist—even in the face of the huge quantities of idle plant and equipment of the greatest depression in history, and the contrary contentions of the "secular stagnationists" who dominated economic thinking in the decade after the *General Theory*—that in the real world there is no tendency toward a diminishing return to capital, so that from the long-run viewpoint the demand for capital is infinitely elastic. Accordingly he rejected as unrealistic the classical notion of a long-run equilibrium quantity of capital that would characterize a stationary state to which the normal development of the economy would allegedly lead (Knight, 1936, 1944a).

These conclusions resulted from Knight's fundamental contention that the very process of increasing the stock of capital must of necessity change the "given conditions" which are supposed to generate diminishing returns. This contention was connected, first of all, with Knight's inclusive concept of capital as encompassing all factors of production, so that capital formation in the real world would lead to the augmentation of all productive agents, thus leaving no fixed factor which would be the source of diminishing returns.

Secondly, stressed Knight, ". . . it is practically impossible to imagine any investment activity in the real world which is not in some degree rationally experimental, in the sense of being reasonably expected to lead to new knowledge having some enduring economic significance. That is, all investment consists, in part, of investment in new knowledge." On the other hand, ". . . we can hardly think of new knowledge being applied without considerable accompanying expenditure on new instruments. This is even more unrealistic than the converse case, previously discussed" (1944a, pp. 40–42).

In brief, Knight's "growth model" was one in which all investment generates "learning by doing," and all technological change is "capital-embodied."

Finally, another factor acting against diminishing returns was the fact that wants did not remain constant, but were changing. "Moreover, it is a familiar fact that in real life much investment actually takes the form of creating or changing consumers' wants and that the result appears in the capital and profit-and-loss accounts of an enterprise" (Knight, 1944a, p. 37).

As an aside, I might conjecture that we can find in Knight's capital theory one of the main sources of his objections (on theoretical, as distinct from political, grounds) to Keynes' *General Theory*. Knight could not accept the implications of this book that the rate of interest was related to the flow of current investment, and not (as Knight had emphasized) the existing stock of capital. Nor could Knight entertain the notion of a long-run stagnating economy in which the rate of interest was driven down to zero. Finally, Keynes' presentation of the rate of interest as a monetary, and not a real, phenomenon was anathema to Knight (1937, pp. 112–13; 1941, p. 221–23; see also 1960, pp. 81, 92).

In any event, Knight's continued emphasis on the dynamic and progressive aspects of the real world—and hence the inappropriateness of analyzing it in terms of a static model—was one of the most important messages that he conveyed to his students. As already noted, he conveyed this message not only with respect to the conditions of production, but also with respect to those of demand. Thus Knight never tired of emphasizing that wants did not remain constant, but constantly expanded over time; that the concept of "minimum" or "subsistence needs" was not a physiological one, but the value judgment of a society at a given point of time as to what the minimum standard of living of its members should be; and that, accordingly, this minimum increased over time with the society's general standard of living. Correspondingly, emphasized Knight, the theory that wages were exogenously determined by the subsistence needs of workers was nonsense.

In this context Knight also contrasted the view on the adjustment of wants to available resources in Western culture, as contrasted with ascetic cultures, such as the Buddhist one: the former attempts to make the adjustment by increasing resources; the latter, by denying wants. And I seem to recall Knight's commenting that from the long-run viewpoint, this denial of wants was the only way that a definitive adjustment of wants to resources could be achieved: for history had shown that Western society created new wants just as fast as (if not faster than!) it expanded the means of satisfying them.

The social determination of wants was a theme that Knight also emphasized in his joint (with Charner M. Perry of the philosophy department) seminar on "Economics and Social Institutions" (Econ. 305). In this sense, then, Knight was an institutionalist. On the other hand, he was not an institutionalist in the sense of advocating the historical or statistical approaches to economics. On the contrary, as I indicated above, Knight's approach to economics was definitely non- (if not anti-) empirical. Indeed, I have a vague recollection of Knight (though it might have been Viner) referring to the excellent fits that have been achieved with the Cobb-Douglas function—and wryly adding that nobody really understood why such consistently good findings emerged from such variegated data![12] In a broader sense of the term, however, Knight was much concerned with empirical facts: and I am referring once again to Knight's insistence

[12] Still, one of the few instances in his writings in which Knight cited a statistical estimate of an economic magnitude was his citation of the Cobb-Douglas estimate of the elasticity of demand for capital! (1936, p. 623).

on basing his analysis of investment and of the wants that lie behind demand on what he discerned as the salient dynamic features of the real world.

These were some of the distinctive features of Knight's teaching of economic theory. Actually, however, he also conveyed these messages in his teaching of the "History of Economic Thought" (Econ. 302), where his detailed critique of the Ricardian system (1935a) was the vehicle by which Knight made clear the main features of his own analytical framework. For this reason this history was live and current for us as students. Here too, however, I must admit that part of this liveliness stemmed from our realizing that the evaluation of Ricardo was also the subject of a running debate between Knight and Viner, where in contrast with the analytical genius whom Viner depicted, Knight depicted a Ricardo who represented a retrogression from the achievements of Smith: primarily I think, because of what Knight regarded as Ricardo's greater emphasis on—and formalization of—the labor theory of value. Indeed, Smith—according to Knight—did *not* believe in the labor theory of value (Classnotes, Econ. 302, 1945).

Knight devoted most of his "History of Thought" course to the classical school, and in this connection had us read the relevant texts from Smith, Ricardo, Senior, and Mill. He then set us the task of presenting what we had learned about this school in a term paper (which had become a standard feature of the course) on "The Classical Cost-of-Production Theory of Price." The remainder of the course was devoted to the Austrian "subjective-value revolution," with a few passing remarks added with reference to the institutionalists. As already indicated, there was much overlapping between this course and Knight's theory course proper—though the emphasis of the latter was on the Marshallian synthesis of the classical and Austrian approaches.

There is one aspect of Knight's treatment of the history of economic thought that continues to puzzle me: despite Knight's broad historical interests—and his 1926 translation of Max Weber's *General Economic History*—Knight made practically no attempt to relate the development of economic thought to the contemporaneous historical developments. At one point in his "History of Economic Thought" course—after distinguishing several "epochs of economic thought"[13]—he went on to say that we "have to examine the historical contexts of the economic writings of these periods" (Classnotes, Econ. 302, 1945). But Knight himself did little further along these lines—except for some passing remarks in his discussion of Ricardo's work.[14] Instead, Knight's approach to the history of thought was that of providing a purely logical criticism of the nature of the assumption made by the various schools of thought, and the validity of the conclusions that they drew from them; he was concerned almost solely with the logical

[13] Namely: "I. Tribal Society and Ancient Empires—in which we do not know how much economic thought went on; II. Greece and Rome; III. Middle Ages; IV. Economic nationalism: Mercantilism (16th–18th centuries); and V. Individualism-liberalism" (Classnotes, Econ. 302, 1945).

[14] "Ricardo in the main was not interested in equilibrium theory, but in the reasons for historical changes: especially in the pressure of increasing population on increasing wealth and constant natural resources. Ricardo was interested in what happens to wages . . . If you project yourself to England of 1815 you can justify this somewhat in terms of what happened during Ricardo's life. About 1770 England began to export manufactured goods and import raw produce—whereas the converse was true before. There were great technological advances [in England]. Then the war cut off importation into England, and so had great increase in agricultural pressure on land. Ricardo was really thinking of revolutionary changes in relative value of agricultural and industrial products" (Classnotes, Econ. 302, 1945).

consistency of the theories he was examining.

At the beginning of this memoir I referred to Knight's humility—and he was indeed humble with respect to the limits of knowledge in general, and of economics in particular. ("I once asked a historian of medicine when he thought doctors began curing more people than they killed. 'Well' he answered, 'I think that that will be in another generation or so'.") But when it came to things that it was given to mortals to understand, Knight (as implied by my description here of his many disagreements with other economists) was a vigorous critic and polemicist—though more so in print and in formal lectures than in his classroom discussions. And even in those cases where it was Frank Knight against the field, we students were left in no doubt as to where the truth lay: "You can be with the majority—or you can be in the right."

II. Knight the Radical Economist—and the Conservative

The general image of Frank Knight in the profession is of a conservative economist whose outstanding students over the years (for example, Henry Simons in the 1920's, Milton Friedman and George Stigler in the 1930's, and James Buchanan in the 1940's) were and are among the leading representatives of this position. So it seems odd at first to think of Frank Knight as a radical economist—but that is one of the sharpest impressions that I have of him.

Knight was not, of course, a radical in the sense of being an advocate of Marxian economics: he had little patience with—and, indeed, had intellectual contempt for—the labor theory of value. His criticism of this theory was expressed in a general way in the course of his analysis of Ricardian economics, mentioned in the preceding section. But he also dealt in a direct—and pithy—manner with some of the basic Marxist tenets. And here Knight revealed one of his major strengths: his ability to make telling points by resorting to obvious facts—whose implications, however, were not at all obvious until he drew them out.

Thus, said Knight, it was "nonsense" to say that labor created capital: "Capital was produced by *capital and* labor working together. Capital is as old as labor" (Classnotes, Econ. 302, 1945, emphasis in original).

Similarly fallacious was the subsistence theory of wages. First of all (as noted above) "subsistence" was itself a changing social concept affected by economic circumstances, and not a fixed physiological concept that could exogenously determine the level of wages. Second, if an employer were indeed free to determine wages at any level he chose, then ". . . an employer will not pay a free laborer a wage high enough to enable him to raise children, since there is no guarantee for the employer to gain by it" (Classnotes, Econ. 302, 1945).

But for the modern economist Knight was—or, more accurately, could have been—a far more effective radical than Marx: for in contrast to Marx, Knight understood the workings of the market system, but he went on by a deeper analysis of these workings to deny the ethical foundations of this system. Indeed, there is little in the writings of the modern day "radical economists" as described, for example, by Lindbeck, that was not more trenchantly said by Knight in his famous essay—published exactly fifty years ago—on "The Ethics of Competition" (1923, especially pp. 45–58). To my mind, these thirteen pages are among the most radical ever written in economics, and even now they remain sharply imprinted in my memory as among the most exciting things that I read during my student days.

Knight's major point in these pages was

that "... in the conditions of real life no possible social order based upon a *laissez-faire* policy can justify the familiar ethical conclusions of apologetic economics" (1923, p. 49).

The reasons for this were, first of all, that

> ... the freest individual, the unencumbered male in the prime of life, is in no real sense an ultimate unit or social datum. He is in large measure a product of the economic system, which is a fundamental part of the cultural environment that has formed his desires and needs, given him whatever marketable productive capacities he has, and which largely controls his opportunities.... It is plainly contrary to fact to treat the individual as a datum, and it must be conceded that the lines along which a competitive economic order tends to form character are often far from being ethically ideal. [1923, pp. 49–50]

Hence (said Knight) it cannot serve as a justification of the market system simply to say that it produces goods that satisfy the wants of individuals in the economy; for a far more fundamental test of any system is the nature of the wants themselves that it produces. " 'Giving the public what it wants' usually means corrupting popular taste" (1923, p. 57).

Or, again, the statement that the market system "works efficiently" is a normative statement, and not a descriptive one: for, said Knight (making use, as he was wont to do, of homely analogies from the world of physics) by the law of conservation of energy, total output is always exactly equal to total input; hence "efficiency" has meaning only as a measure of *useful* output to input, which means that a value judgment must be made as to the "usefulness" of the output to society. The market system "gets things done"; yes, said Knight, but what things?

So it should not surprise us that long before the recently renewed emphasis on the limitations of "consumer sovereignty" (see, for example, John Kenneth Galbraith, pp. 202–28), Knight was emphasizing that wants are not only the product of the system in general, but can also be specifically influenced by advertising:

> ... skill in verbal utterance, or capacity in any form to "influence" other persons, is a form of economic power. Men have lived by their wits from time immemorial; it was no doubt one of the earliest uses of language. Persuasive power is, in the first place, a form of "productive capacity" in that skilful "puffing" makes a product more desirable to a purchaser and consumer, who in an individualistic system is the final judge of the merits of products.... Adam Smith and the classical economists, in advocating freedom of access to the market, do not seem to have thought about the possibilities of personal influence as a serious factor in economic relations; advertising and selling technique assumed importance after Smith's day [1935b, p. 292; see also 1923, p. 51, note]

Another deficiency of the market system is that it tends naturally towards monopoly:

> No error is more egregious than that of confounding freedom with free competition, as is not infrequently done. As elementary theory itself shows, the members of any economic group can always make more by combining than they can by competing. The workings of competition educate men progressively for monopoly, which is being achieved not merely by the "capitalist" producers of more and more commodities, but by labour in many fields, and in many branches of agriculture, while the producers of even the fundamental crops are already aspiring to the goal. [1923, p. 52]

(I should note that in later years Knight came to consider labor unions—with their growing power—as the most harmful form of monopoly in the economy (see for example (1960, pp. 168–70).)

Knight's sharpest shafts were reserved for his critique of "productivity ethics"—

the view (that Knight usually associated with the name of J. B. Clark) that one could attribute an ethical value to the income distribution generated by paying factors of production in accordance with their respective marginal products.

First of all, said Knight, because of market imperfections in the real world, "there is only a 'general tendency' to impute to each productive agent its true product" (1923, p. 55).

Second, the value of the marginal product depends on its price; and the

> money value of a product is a matter of the "demand," which in turn reflects the tastes and purchasing power of the buying public and the availability of substitute commodities. All these factors are largely created and controlled by the workings of the economic system itself, as already pointed out. Hence their results can have in themselves no ethical significance as standards for judging the system. [1923, p. 55]

And to these two points, Knight added his most trenchant criticism of all:

> The income does not go to "factors," but to their owners, and can in no case have more ethical justification than has the fact of ownership. The ownership of personal or material productive capacity is based upon a complex mixture of inheritance, luck, and effort, probably in that order of relative importance. What is the ideal distribution from the standpoint of absolute ethics may be disputed, but of the three considerations named certainly none but the effort can have ethical validity. [1923, p. 56]

Not only is there no ethical justification for the existing distribution of income (and, as can be seen from the preceding citation, Knight did not distinguish in this context between income from labor and income from property), but as a result of the cumulative processes at work in the market system, the distribution becomes increasingly worse:

> . . . where the family is the social unit, the inheritance of wealth, culture, educational advantages, and economic opportunities tend toward the progressive increase of inequality, with bad results for personality at both ends of the scale. [1923, p. 50]

These words were written in one of Knight's first articles, when he was a relatively young man (around 35). And I am sure that there will be those who will say that they represent the "radicalism of youth" which Knight later "outgrew."

To this I can only say (on the basis of my own classroom experiences) that even twenty years later Knight continued to consider these criticisms sufficiently important to convey in his lectures to his students.[15] Furthermore, he had earlier repeated many of them in his major essay on "The Ethics of Liberalism" (1939a, pp. 45–74). I also find it significant that—at the age of 65—he returned to these points in what was essentially a statement of his credo in his presidential address to the American Economic Association (1951). Similarly, in one of the last articles he wrote (1966),[16] Knight made use of the foregoing points (and others) to present a sharp criticism of a would-be justification for the capitalist system by Henry Hazlitt[17,18] Thus the evidence indicates

[15] These discussions recurred in all of Knight's courses. They were, however, particularly emphasized in a joint seminar (again, with Charner M. Perry) on "Economic Theory and Social Policy" (Econ. 304).

[16] I am indebted to Paul Samuelson for bringing this article to my attention.

[17] Several readers of the first draft of this memoir have called my attention to a lecture before a student group which Knight gave during the presidential elections of 1932 and which was entitled "The Case for Communism from the Standpoint of an Ex-Liberal," a lecture of whose existence I was not hitherto aware. The main burden of this lecture, however, was not Knight's ethical critique of the economic functioning of the capitalist system that has just been described, but a scathing criticism of the inability of a democracy to govern itself rationally and efficiently by means of discussion—from which Knight concluded that government by a communist dictatorship might be preferable. This emphasis is also clear from the general title under which this lecture—as well as two others that Knight

that Knight's fundamental ethical critique of the actual operation of the market economy was an essential part of his thinking throughout his life.

And so at the Chicago of my student days it was, ironically enough, the socialist Oskar Lange who extolled the beauties of the Paretian optimum achieved by a perfectly competitive market—and Frank Knight who in effect taught us that the deeper welfare implications of this optimum were indeed quite limited.

All of which raises the question as to the sharp contrast between Knight's incisive criticism of the existing system, and his basically conservative opposition to various proposals made to reform it, and all the more so to change it radically. For this conservatism, too, was a fundamental element of Knight's teachings. Thus after summarizing his ethical critique of the market system in his 1939 "Ethics of Liberalism," Knight went on to conclude that

> . . . society has strong reasons for maintaining powerful brakes on departures from the "beaten path." Primitive society was wise in its conservatism, for it knew at least that the group had previously lived somehow, both as individuals and as a group. And liberal society, it now seems, has acted frivolously in switching over quite suddenly to an extreme opposite set of assumptions, that the new is better than the old, that the good consists in change, or at least in freedom of the individual to make changes; rather than in stability. This emphasis on the necessity of an *onus probandi* in favour of conservatism and against change, must stand as our last word at this point. [1939a, p. 74; see also 1935b, p. 286]

Similarly, in his 1950 presidential address he emphasized that ". . . the possible amount and speed of free and intelligent social change will always be quite narrowly limited" (1951, p. 276).

In some cases Knight's opposition to reform proposals stemmed from his belief that the proposals in question failed to reveal a proper understanding of market forces. Thus he opposed price and rent controls because he felt that these reflected a failure to understand the functioning of the price system in allocating scarce resources. He was particularly critical of reform proposals that in his eyes committed the original sin of offering "something for nothing,"[19] and he continuously decried the age-old human fail-

gave before student groups at roughly the same time—were subsequently multilithed for private circulation: namely, *The Dilemma of Liberalism* (1933b). I might also note that Knight included the main substantive points of this criticism of the democratic process in his essay a few years later on "Economic Theory and Nationalism" (1935b), discussed below.

There is no copy of the *Dilemma of Liberalism* in the University of Chicago library, though there is a copy (to which I was directed by Maynard Kreuger) in the library of the London School of Economics. A copy (though without the title page) is also to be found among the papers that Knight left after him.

I might also note that the sections of the lecture on "The Case for Communism from the Standpoint of an Ex-Liberal" dealing with the genesis, character, and failures of the liberal regime (1933b, pp. 21–45) were reprinted in the eighth (1939) edition of the *Selected Readings* for the Social Science II survey course (Knight, 1939b). However, the title of the lecture from which these sections were taken (namely, "The Case for Communism . . .") is blackened out both in the Table of Contents of the *Selected Readings* and at the head of the reprinted sections, though it remains partly legible. A footnote attached to the blackened-out title indicates that the selection reprinted is an "excerpt from a privately printed pamphlet."

[18] I should, however, note the milder tone of some of these criticisms in Knight's 1960 book on *Intelligence and Democratic Action*, written as he approached the age of 75. Here Knight also wrote:

> I may say that in the three decades or so since I laboriously worked up a lecture on "The Ethics of Competition" (given at Harvard University and later published as an article in the *Quarterly Journal of Economics*—and still later reprinted in a book) I have done quite a lot of thinking about ethics and economics, and have perpetrated some wordage in print as well as in several university classrooms. As a result, I have become even more hesitant about speaking very definitely and positively. [1960, p. 122]

[19] From which view on life may have stemmed the more recent "Chicago tradition" that "there is no such thing as a free lunch."

ings that led to the belief in such a possibility. "A refrigerator generates cold," he would say in another one of his parables from physics, "only by conveying heat from the inside to some point on the outside. And people just never seem to be able to understand that simple fact."

But a far more basic source of Knight's conservatism was his view that as bad as the existing market system was, the alternatives were even worse! Indeed, this is the view to which he adverts in the concluding paragraph of his discussion in the aforementioned "thirteen pages" of his "Ethics of Competition":

> It is expressly excluded from the field of the present paper to pass any practical judgment upon the competitive system in comparison with any possible alternative. But in view of the negative tone of the discussion, it seems fair to remark that many of the evils and causes of trouble are inherent in all large-scale organization as such, irrespective of its form. It must be said also that radical critics of competition as a general basis of the economic order generally underestimate egregiously the danger of doing vastly worse. Finally, let us repeat that practically there is no question of the exclusive use or entire abolition of any of the fundamental methods of social organization, individualistic or socialistic. Economic and other activities will always be organized in all possible ways, and the problem is to find the right proportions between individualism and socialism and the various varieties of each, and to use each in its proper place. [1923, p. 58]

To the best of my knowledge, Knight never specified the criteria by which these "right proportions" were to be chosen—and I shall return to this point below. What I want now to explain is that the way in which Knight feared that the alternatives could ethically do "vastly worse" than the existing market system was primarily in their impact on individual freedom. For Knight, such freedom was an absolute value: even if people do not want freedom, they ought to want it. Furthermore (and here Knight's historical perspective reinforced his conservatism) liberalism and individual freedom were recent historical developments.[20] Indeed, they might even be said to have been the chance product of "a remarkable and temporary concourse of circumstances" that prevailed during the eighteenth and nineteenth centuries. Thus individual freedom was a fragile development—in constant danger of being destroyed (1935b, pp. 287–89, p. 305, note; see also 1960, pp. 38–39, 108).

"There has been, and must be, a question," said Knight in his critical review of Sumner Slichter's book in favor of economic planning[21]

> . . how far it is *comparatively* better social policy (I am aware of the redundancy but wish to be clear) to have individuals make their own choices in economic life and how far better to have them dictated—by anybody else who can be thought of in the position of dictating them. . . . What I want from the preachers of control, and do not get in [Slichter's] *Modern Economic Society* . . . or anywhere else, is something I can understand on the subject of *who* is to make the economic choices for the individual who admittedly makes them quite "imperfectly," or *how* this choosing functionary is to be selected, and *how* the individual affected is to be brought to accept them; and, in general, what kind of (a) individuals and (b) social order the preachers are either assuming or working toward. [1932b, pp. 823–24, italics in original]

Looking back on all this now, I would conjecture that an important additional element of Knight's fundamental conservatism—and general opposition to "planning" or other government action

[20] The Greek city-states with their slaves were, in Knight's eyes, hardly in this tradition.

[21] I am indebted to George Stigler for bringing this review article to my attention.

whose declared aim was the improvement of the actual workings of the market economy—was his view of the "system of individual freedom" as representing a stage of unstable equilibrium in the course of human history. (I think it is significant that the instability of various aspects of the macro-world is a recurrent element of Knight's thought: it is reflected, as we have seen, in his discussion of the cumulative tendencies toward inequality and monopoly in a market economy; it is also reflected in his discussion (to which I shall now turn) of the cumulative tendency toward the concentration of power; finally, in a different context to be discussed below, it is reflected in his view of the business cycle—and particularly of the cumulative movement in it of the absolute price level.) In view of the existence of this knife-edge equilibrium, Knight feared that any encroachment by the state upon individual freedom in the attempt to "improve" the economy would generate cumulative departures from this freedom.

Be that as it may, the specific ways in which he feared that one of the possible alternatives to the market system—"democratic socialism"[22]—could "do worse" were spelled out by Knight in his lengthy 1935 essay on "Economic Theory and Nationalism." Knight's point of departure was that socialism was the extension of the political process to the business world. But if this extension were carried out within a democratic framework (as its advocates claimed to be interested in doing) then

> . . . the essential point is that, as it has worked out in practice in the modern world, *democracy is competitive politics*, somewhat as free enterprise is competitive economics (though inherently a competition for a monopolistic position), and shows the same weaknesses as the latter. [1935b, p. 295, italics in original]

In particular, competitive politics cannot be described as a system devoted to satisfying the given wants of the public; for this system—even more than competitive economics—tries by "advertising" (i.e., campaigning and political propaganda in general), to influence and form these wants themselves. Or, as Knight had succinctly expressed himself a few years earlier in sharp criticism of Sumner Slichter's defence of planning, ". . . between business and politics, it is as easy and as reasonable to assert that corruption goes one way as the other. . . . Both finally 'give the people what they want,' after doing their utmost to make them want what they want to give" (1932, p. 476). And, once again, the nature of the wants so generated by competitive politics are of questionable ethical value: ". . . in politics the appeal is almost exclusively to the crowd" (1935b, p. 299). Similarly questionable are the human qualities promoted by competitive politics as a means of succeeding in them.

Finally, the system of competitive politics that characterizes democracy and "democratic socialism" in particular will tend to generate more inequality in the real sense of the term than does the system of competitive economics. For ". . . the significance of consumption itself is largely symbolic; the inequality which really hurts is the unequal distribution of dignity, prestige, and power" (1935b, pp. 308–309; see also p. 298, note). And to this must be added the fact that the abilities which make for success in competitive

[22] By which Knight meant a representative democracy that continued to make use of the market, but in which

> . . . entrepreneurs and property owners . . . would be replaced by administrative officials, who, like their prototypes, would theoretically have no discretionary power over production, the details of organization being determined, just as under the theory of enterprise, by consumers' choices. [1935b, p. 307]

Presumably, Knight had in mind the kind of socialist system that Fred Taylor had described in his 1928 AEA presidential address, and which Lange (1936, 1937) and Lerner (1937) were later to explicate in greater detail.

politics ". . . are more unequally distributed among men by nature than is economic ability or power of any other kind, and also tend more strongly to cumulative increase through their own exercise" (1935b, pp. 296–97).

Thus not only does power corrupt, but it corrupts cumulatively. Nor can free elections be considered as a safeguard against this cumulative concentration of power in the hands of political leaders:

> In view of the way in which psychological principles work in politics, equal suffrage (even if it is respected in practice) provides little or no guarantee of equality in that field. The overwhelming majorities rolled up in plebiscites (more or less fairly conducted) on the question of dictatorship show where the realities lie. [1935b, p. 297; see also pp. 308–09]

(It should be recalled that Knight was writing these words a few years after Hitler had democratically reached power in Germany—as had Mussolini before him in Italy.)

The depth and intensity of Knight's feelings on the issues discussed in the preceding pages are well represented in the following paragraph from Knight's presidential address—or his "sermon," as he denoted it:

> The plea of communism, like that of Christianity, is justice, under absolute authority, ignoring freedom. (The former does extol progress, and progress through science, both of which Christianity despised; by the same argument, communism is overtly less devoted to law and tradition, more openly claims the right to ignore or break the law.) For liberalism, the primary value is freedom, self-limited by laws made by the community, ideally by general assent, in practice by representatives elected by a voting majority—one of its dangers. The laws of a liberal state will also be general, non-specific, but in a sense quite different from the Golden Rule or Law of Universal Love. The familiar figure is "rules of the road," in contrast with instructions where and when and how to travel, whether arbitrary or conformable to a traditional practice. But such freedom must be sweepingly limited by measures, not only of a "police" character in a broad sense, but also designed to equip the individual and family for social life by implanting wants and tastes in general conformity with the culture and endowing each with a minimum of productive capacity (or ultimately with final goods) without which freedom is a form empty of content. To take these units as "given' is flagrantly contrary to essential facts of life and means ignoring the major social problem. It is along this line that eighteenth- and nineteenth-century liberalism went to an extreme that has provoked a reaction which threatens to engulf all freedom, and justice too, in the modern conception of it, if not to destroy civilization. Liberal states have been engaged, however, through their short life, in correcting this imbalance between freedom and justice; more intelligence and better judgment is our need, rather than any radical departure in method. [1951, pp. 277–78]

In concluding this discussion of Knight's political views, I would like to turn to a question that arises most naturally from it: what was the relationship between Knight and the "Chicago school"? From one viewpoint this is a rhetorical question: for in his emphasis on individual freedom as a basic value, Knight (together with Viner) can only be considered to be one of the "founding fathers" of this school. At the same time I think that Knight (as well as Viner) was less doctrinaire in his views than some of his younger colleagues and followers. And clearly Knight (and here he differed from Viner) did not have the policy-orientation that is one of the hallmarks of the Chicago school.[23]

[23] This should not be considered to be simply a reflection of the absence of an empirical orientation on the part of Knight. For Henry Simons and Lloyd Mints were not empirically oriented either.

Thus, as we have seen, Knight was sharply critical of the ethical deficiencies of the actual workings of the market system—and was particularly scornful of the "productivity ethics" by which this system was sometimes justified. Nor did he make an absolute distinction between "government by law" and "government by men"—for laws too are made by men. In particular, Knight emphasized that

> . . . the primary difficulty with the notion of law as an ethical principle or norm is that the content of the law itself can never be taken as simply "given," or beyond dispute, even at a given moment. . . . there was always and inevitably occasion for "interpreting" the law, in enforcing it, and also for making law outright, i.e., changing it, in consequence of changing conditions and standards. [1939a, pp. 62–63]

A similar picture holds with respect to the specific monetary policy with which the Chicago school of the 1930's and early 1940's is so strongly identified: namely, the contracyclical policy of adjusting the quantity of money so as to stabilize the price level. The rationale for this "monetary rule" had been developed most systematically during the 1930's by Henry Simons,[24] who continued to expound it in his courses and seminars with us on fiscal policy. And Lloyd Mints did the same thing in his courses on monetary theory and policy.

In contrast, as I have already shown in a paper on the monetary policy of the Chicago school,[25] though Knight recognized the importance of monetary disturbances in generating cycles of booms and depressions, he was skeptical about the extent to which these cycles could be controlled. As he expressed himself in one of his lectures:

> In medieval times men didn't look for remedies since they thought everything was from God who was good—so everything was good. Now science is the God—and we think that there must be a remedy for every disease. Maybe there is no answer to the business cycle: maybe we have to let it take its course. [Classnotes, Econ. 301, 1945]

Correspondingly, I think it fair to say that Knight was less convinced than his younger colleagues, Simons and Mints, about the efficacy of the monetary rule of stabilizing the price level as a means of combatting the cycle. Indeed, in one of his lectures Knight said that he did "not know how to stabilize the price level—and at what height to stabilize it" (Classnotes, Econ. 301, 1945). I might also note that on at least one occasion in later years Knight expressed his preference for "administrative discretion" rather than "rules" in the carrying out of monetary policy (1960, pp. 105–06).[26]

But in all this—as well as in his failure noted above to specify the criteria by which to decide upon "the right proportions between individualism and socialism"—I think that we must see the manifestation of yet another basic element of Knight's character: that in what he regarded as the inherent contradiction between thought and action, Knight remained forever the man of thought: always critical, always probing, always asking questions—and always deeply aware of the complexities and inertias of the real world, and hence skeptical about the ex-

[24] Simons had been a student of Knight's at the University of Iowa, and had followed him to Chicago in 1927 as a lecturer.

[25] (1969, pp. 96, note 13; 111–12; and 114–15.) In 1969, I sent a copy of this paper to Knight and asked him for his comments. But Knight replied that "to be quite frank, I'm very definitely 'retired' from economics"—and went on to write briefly about other matters, including his concern about the tensions in the Middle East. (Personal letter to author, dated March 19, 1969.)

[26] It might be noted that Viner (1962), too, expressed his preference for discretionary action on the part of the monetary authorities. On the general relation of Viner to the Chicago school, see Patinkin (1969, pp. 94, note 5; 96, note 13; and 113.)

tent to which man could by deliberate policy-actions improve it.[27]

III. Knight the Social Philosopher and the Man

In my introductory reminiscences, I mentioned that I had heard Knight's "Price and Distribution Theory" course twice—and there is a distinct difference between my classnotes from these two exposures: for whereas the first is filled primarily with economic matters proper, the second contains less of these, and more of long digressions on the nature of man and society—and God.

But I do not think that this difference between the classnotes necessarily reflects a difference in the actual emphasis of these two courses. Instead, I suspect that it reflects the difference that had taken place in me, the listener. For by the second time I took the course I had already had several courses in economic theory proper, and I had also come to understand that the unique contribution that Knight as a teacher had to make to students—and not only to students!—was precisely in his rambling "digressions" on the nature of the world.

One of his major concerns in these digressions was religion—a concern that undoubtedly stemmed from Knight's growing up in the deeply religious atmosphere of rural Illinois at the turn of the century. After his youth, Knight was fighting religion all his life, and his strongest and most violent language were reserved for expressing his anticlerical views.

There were several dimensions to these views. First, the irrationality of religion—and even more specifically, its ban on intellectual curiosity and the asking of questions, its insistence on accepting its tenets without further thought, All this was anathema to Frank Knight—the eternal asker of questions ("Do parents in primitive societies correct their children's grammar?").

But I think that an even more important cause of Knight's anticlericalism was the discrepancy between the declarations of organized religions and their behavior. Again in the words of his presidential address:

> In Christianity, surely, we find the supreme "irony of history": that an original teaching centered ethically in humility, meekness, self-denial, and self-sacrifice became organized into corporations whose dignitaries have hardly been matched for arrogant grasping, using, and flaunting of power and wealth and for insistence on prerogative to the borderline of worship. One turns to Dostoevski's famous speech of the Grand Inquisitor for any adequate portrayal of this situation and its sinister indications of the nature of human nature. [1951, p. 277]

I believe that it was because of his confrontation with this discrepancy in his own cultural heritage that Knight's anticlericalism expressed itself in anti-Christianity, and anti-Catholicism in particular. I feel (though I may be prejudiced) that he expressed less opposition to Judaism: not because he thought any better of Judaism as a religion, but because the historical development of Judaism had been such as to leave it (in Knight's life experience) without the institutional frame-

[27] The following words from Knight's last book, written as he approached the age of 75, are revealing:

> It has been said that fools ask questions and wise men answer them . . . To that indictment I must plead guilty, and also that of being better at criticizing other people's asking and answering than at doing either myself, and more addicted to the former role. But I have a defense. . . . I think the role fits in with the concept of democracy, that the function of the intellectual leader, in the difficult field of social philosophy and policy, is to clarify issues, or at most suggest possible solutions and perhaps arguments, pro and con. Answering is rather the task of Adam Smith's insidious and wily animal, the statesman or politician, which in a democracy means the ordinary citizen. [1960, p. 121]

work and consequent power that so aroused his wrath in the case of Christianity.

In this context it is also significant that (to the best of my knowledge) Knight did not criticize Buddhism or Islam or any other of the great world religions in the same way that he did Christianity: and, I suspect, for the same reason just indicated—namely, that he had not directly experienced and been irritated by the workings of their institutional frameworks.

Here too we find the kind of contradiction in Knight's views that I noted above with respect to the question of social reform. He was highly critical of religion—and yet he recognized the vital function that it performed in maintaining the stability of any society: because the possibilities of social reform were limited—and could take place only slowly without endangering the existence of society—"men will always require—as a condition for maintaining any high civilization at all, some 'opiate' or some effective agent to prevent their demanding their rights" (1951, p. 276).

But if Knight was not a religious person in those things that concern man and God, in the things that concern man and man he exemplified the ethical qualities of moral rectitude and personal integrity that all religions teach us to aspire to. And he was an incessant and unswerving Seeker after the Truth.

I have mentioned Knight's opposition to the irrationality of religion. At the same time, however, he was the very opposite of those who turned rationality itself into a religion. Indeed, for Knight the very desire for rationality was one of the expressions of man's romantic nature. Knight never tired of quoting the saying—that he attributed to J. M. Clark (p. 24)—about the "irrational passion for dispassionate rationality." And he always reminded us that it was not the people of low intelligence who spent their lives squaring the circle and inventing perpetual-motion machines.

Similarly, as already noted, he was deeply skeptical of the extent to which rational discussion could solve social problems: in part, because of the intrinsic difficulty of these problems themselves; but even more so because of the difficulty of reaching a consensus; and the more intelligent people were, the finer the distinctions they would insist on making, and hence the less likely they were to achieve such a consensus. And perhaps most important of all: free discussion in a democracy would not remain rational, but would degenerate into demagoguery: "Cheaper talk drives out of circulation that which is less cheap"—so "Frank Knight's First Law of Talk" (1933b, p. 8).

Indeed, here we find the deepest contradiction in Knight's view of human society: on the one hand, he regarded individual freedom as a basic value, and recognized that representative democracy was the only way in which a large society of free individuals could govern itself; on the other, he had basic misgivings about the actual workings of the democratic process—and was accordingly deeply pessimistic about its future. So much so that on some occasions he predicted the "natural" disintegration of democracy and its replacement by dictatorship.

Another of Knight's major themes (and an important component of his anti-religious position) was that it was not enough to want to do good: one had to have an understanding of the workings of social forces in order to be able actually to do good. The church might outlaw usury as unjust: but because of the workings of market forces all that this prohibition did was to change the form in which income from property was received—from the form of interest to that of rents or profits.

The government might—in the face of "shortages" of critical goods—impose price controls: but this would not increase the quantities of these goods, and would indeed create artificial scarcities.

And "love thy neighbor as thyself" was not a prescription for social action. How do you know that your neighbor has the same tastes as you do? More important—to love everybody is to love nobody: for love is a differential relationship.[28]

Knight was deeply concerned with the personal element of morality and the ethical problems created by an impersonal society. No one (Knight once said) would reinsert his coin into a public phone if it were accidentally returned to him after the completion of his call; indeed, if he were to do so, he would be considered a fool—not a saint. But was not this stealing?

Cynical? Perhaps. Though not the cynicism of an embittered man, but the cynicism of one who looked with understanding—and, though he would probably have denied it, even compassion—on the limitations of man and society. The cynicism of one who had resigned himself to the limitations of *la condition humaine*.

He was—as has been said of Ecclesiastes—a gentle cynic.

REFERENCES

G. C. **Archibald**, "Chamberlin Versus Chicago," *Rev. Econ. Stud.*, Oct. 1961, *29*, 1–28.

———, "Reply to Chicago," *Rev. Econ. Stud.*, Feb. 1963, *30*, 68–71.

J. M. **Buchanan**, "Frank H. Knight," in *International Encyclopedia of the Social Sciences*, New York 1968, Vol. 8, 424–28.

E. H. **Chamberlin**, *The Theory of Monopolistic Competition*, 4th ed., Cambridge, Mass. 1942.

[28] See Knight (1935b, p. 312; 1939a, pp. 102–28; and 1945).

J. B Clark, *The Distribution of Wealth*, New York 1899.

J. M. Clark, "Economics and Modern Psychology," Part I, II, *J. Polit. Econ.*, Jan. 1918, *26*, 1–30; Feb. 1918, *26*, 136–66.

H. J. Davenport, *The Economics of Enterprise*, New York 1913.

J. Dorfman, *The Economic Mind in American Civilization*, Vol. 5, New York 1959.

C. E. Ferguson, *Microeconomic Theory*, rev. ed., Homewood 1969.

M. Friedman, "The Marshallian Demand Curve," *J. Polit. Econ.*, Dec. 1949, *57*, 463–95; reprinted in *Essays in Positive Economics*, Chicago 1953, 47–99.

———, "More on Archibald Versus Chicago," *Rev. Econ. Stud.*, Feb. 1963, *30*, 65–67.

J. K. Galbraith, *The New Industrial State*, Boston 1967.

Y. **Grunfeld**, N. **Liviatan**, and D. **Patinkin**, *Lectures on Price Theory: Winter Term 1959/60*, (Hebrew), mimeo., Jerusalem 1963.

T. Haavelmo, *A Study in the Theory of Investment*, Chicago 1960.

R. F. Harrod, "Doctrines of Imperfect Competition," *Quart. J. Econ.*, May 1934, *48*, 442–70.

H. Hazlitt, *The Foundations of Morality*, Princeton 1964.

J. R. Hicks, *Value and Capital*, Oxford 1939.

A. Johnson, *Pioneer's Progress*, Omaha 1960.

J. M. Keynes, *The General Theory of Employment Interest and Money*, New York 1936.

F. H. Knight (1921a), *Risk, Uncertainty, and Profit*, Boston 1921, New York 1957.

——— (1921b), "Cost of Production and Price Over Long and Short Periods," *J. Polit. Econ.*, Apr. 1921, *29*, 304–35; reprinted in *The Ethics of Competition and Other Essays*, New York 1935, 186–216.

———, "Economics, 1920's," (Unpublished material in Boxes 8 and 9 of Knight's papers that have been catalogued by Glen Gilchrist under George Stigler's direction.) In part, early versions of 1932c, apparently written during the period 1922–26 (see fn. 5 above).

———, "The Ethics of Competition," *Quart. J. Econ.*, Aug. 1923, *37*, 579–624; reprinted in *The Ethics of Competition and Other Essays*, New York 1935, 41–75.
———, "Herbert Joseph Davenport (1861–1931)," in *The Encyclopedia of the Social Sciences*, New York 1931, Vol. 5, 8–9.
———, (1932a) "Interest," in *The Encyclopedia of the Social Sciences*, New York 1932; reprinted in *The Ethics of Competition and Other Essays*, New York 1935, 251–76.
———, (1932b) "The Newer Economics and the Control of Economic Activity," review of S. H. Slichter, *Modern Economic Society*, *J. Polit. Econ.*, Aug. 1932, *40*, 433–76.
———, (1932c) "Modern Economic Society Further Considered," *J. Polit. Econ.*, Dec. 1932, *40*, 820–25.
———, (1932d) "Social Economic Organization," "The Price System and the Economic Process," "Demand and Supply and Price," "Distribution: The Pricing of Productive Services Individually," in *Second Year Course in the Social Sciences (Social Sciences II): Syllabus and Selected Readings*, preliminary ed., multilith, Chicago 1932.
———, (1933a) *The Economic Organization*, Chicago 1933, multilith; reprinted New York 1951. (Reproduces the material of the preceding entry.)
———, (1933b) "The Case for Communism: From the Standpoint of an Ex-Liberal," in *The Dilemma of Liberalism*, Ann Arbor 1933 (multilithed for private circulation).
———, (1935a) "The Ricardian Theory of Production and Distribution," Part I, II, *Can. J. Econ.*, Feb. 1935, *1*, 3–25, May 1935, *1*, 171–96; reprinted in *On the History and Method of Economics*, Chicago 1956, 37–88.
———, (1935b) "Economic Theory and Nationalism," in *The Ethics of Competition and Other Essays*, New York 1935, 277–359.
———, (1935c) *The Ethics of Competition and Other Essays*, New York 1935.
———, "The Quantity of Capital and the Rate of Interest," Part I, II, *J. Polit. Econ.*, Aug. 1936, *44*, 433–63; Oct. 1936, *44*, 612–42.
———, "Unemployment and Mr. Keynes' Revolution in Economic Theory," *Can. J. Econ.*, Feb. 1937, *3*, 100–23.
———, (1939a) "The Ethics of Liberalism," in "Ethics and Economic Reform," *Economica*, Feb. 1939, *6*, 1–29; reprinted in *Freedom and Reform*, New York 1947, 45–74.
———, (1939b) "The Genesis and Character of the Modern Liberal Regime," and "The Breakdown of the Liberal System: Its Weaknesses—Reasons for Failure," in *Selected Readings for the Second-Year Course in the Study of Contemporary Society (Social Science II)*, 8th ed., Chicago Sept. 1939 (multilith, reproduces pp. 21–45 of 1933b).
———, "The Business Cycle, Interest, and Money: A Methodological Approach," *Rev. Econ. Statist.*, May 1941, *23*, 53–67; reprinted in *On the History and Method of Economics*, Chicago 1956, 202–26.
———, (1944a) "Diminishing Returns from Investment," *J. Polit. Econ.*, Mar. 1944, *52*, 26–47.
———, (1944b) "Realism and Relevance in the Theory of Demand," *J. Polit. Econ.*, Dec. 1944, *52*, 289–318.
———, "Liberalism and Christianity," in *The Economic Order and Religion*, New York 1945, 13–126.
———, "Comment on Mr. Bishop's Article" [on the theory of demand], *J. Polit. Econ.*, Apr. 1946, *54*, 170–76.
———, *Freedom and Reform*, New York 1947.
———, "The Rôle of Principles in Economics and Politics," *Amer. Econ. Rev.*, Mar. 1951, *41*, 1–29; reprinted in *On the History and Method of Economics*, Chicago 1956, 251–81.
———, *On the History and Method of Economics*, Chicago 1956.
———, *Intelligence and Democratic Action*, Cambridge, Mass. 1960.
———, "Abstract Economics as Absolute Ethics," *Ethics*, Apr. 1966, *76*, 163–77.
——— and T. W. Merriam, *The Economic Order and Religion*, New York 1945.

O. Lange, "On the Economic Theory of Socialism," Part I, II, *Rev. of Econ. Stud.*, Oct. 1936, *4*, 53–71; Feb. 1937, *4*, 123–42; reprinted in O. Lange and F. M. Taylor, *On the Economic Theory of Socialism*, Minneapolis 1938, 55–143.

———, "The Rate of Interest and the Optimum Propensity to Consume," *Economica*, Feb. 1938, *5*, 12–32; reprinted in G. Haberler, ed., *Readings in Business Cycle Theory*, Philadelphia 1944, 169–92.

———, *Price Flexibility and Employment*, Bloomington 1945.

A. P. Lerner, "Statics and Dynamics in Socialist Economics," *Econ. J.*, June 1937, *47*, 253–70.

———, *The Economics of Control*, New York 1944.

———, "On the Marginal Product of Capital and the Marginal Efficiency of Investment," *J. Polit. Econ.*, Feb. 1953, *61*, 1–14.

A. Lindbeck, *The Political Economy of the New Left—An Outsider's View*, New York 1971.

R. G. Lipsey, *An Introduction to Positive Economics*, 2d. ed., London 1966.

A. Marshall, *Principles of Economics*, 8th ed., London 1920.

H. L. Miller, Jr., "On the 'Chicago School of Economics'," *J. Polit. Econ.*, Feb. 1962, *70*, 64–69.

D. Patinkin, "Demand Curves and Consumer Surplus," in Carl Christ et al., eds., *Measurement in Economics: Studies in Mathematical Economics and Econometrics in Memory of Yehuda Grunfeld*, Stanford 1963, 83–112.

———, "The Chicago Tradition, the Quantity Theory and Friedman," *J. Money, Credit, Banking*, Feb. 1969, *1*, 46–70; reprinted in *Studies in Monetary Economics*, New York 1972, 92–117.

———, "In search of the 'Wheel of Wealth': On the Origins of Frank Knight's Circular-Flow Diagram," *Amer. Econ. Rev.*, Dec. 1973, *63*, 1037–46.

P. A. Samuelson, *Economics: An Introductory Analysis*, New York 1948, 1951 (2d. ed.), 1958 (4th ed.) 1970 (8th ed.).

———, (1972a) "Economics in a Golden Age: A Personal Memoir," in G. Holton, ed., *The Twentieth Century Sciences: Studies in the Biography of Ideas*, New York 1972, 155–70.

———, (1972b) "Jacob Viner, 1892–1970," *J. Polit. Econ.*, Jan./Feb. 1972, *80*, 5–11.

S. H. Slichter, *Modern Economic Society*, New York 1931.

H. W. Spiegel, "Herbert J. Davenport," in *International Encyclopedia of the Social Sciences*, New York 1968, Vol. 4, 16–17.

G. J. Stigler, *The Theory of Price*, rev. ed., New York 1952.

———, "Archibald Versus Chicago," *Rev. Econ. Stud.*, Feb. 1963, *30*, 63–64.

F. M. Taylor, "The Guidance of Production in a Socialist State," *Amer. Econ. Rev.*, Mar. 1929, *19*, 1–8; reprinted in O. Lange and F. M. Taylor, *On the Economic Theory of Socialism*, Minneapolis 1938.

J. Viner, 'Cost Curves and Supply Curves," *Zeitschrift für Nationalökonomie*, Sept. 1931, *3*, 23–46; reprinted in *The Long View and the Short*, Glencoe 1958, 50–78.

———, "The Necessary and the Desirable Range of Discretion to be Allowed to a Monetary Authority," in L. B. Yeager, ed., *In Search of a Monetary Constitution*, Cambridge, Mass. 1962.

M. Weber, *General Economic History*, translated by F. H. Knight, New York 1927.

Labor Hoarding in Durable Goods Industries

By C. Scott Clark*

Most economists are familiar with the observed short-run behavior of productivity as measured by output per man or output per man-hour. It is generally observed that productivity tends to rise most rapidly when output is recovering towards capacity and to fall or rise less rapidly as output declines from capacity. According to the law of diminishing marginal productivity, however, the opposite behavior should be observed. That is, with capital fixed in the short run, there should be decreasing returns to labor as output increases towards capacity and increasing returns to labor as output declines from capacity. This apparent contradiction between reality and theory has not been confined solely to studies of how output fluctuates with employment. Econometric studies of short-run employment behavior have consistently obtained estimates of the equilibrium output-employment elasticity which suggest increasing returns to labor services (see Robert Ball and E. B. A. St. Cyr, Frank Brechling, David Smyth and N. J. Ireland, and Robert Solow).

A number of hypotheses have been put forward in an attempt to reconcile this apparent paradox (see Ray Fair, Edwin Kuh, and Robert Solow). One possible explanation, which recently has begun to attract an increasing amount of support, is referred to as the labor-hoarding theory. This theory operates on the assumption that there are important costs associated with changing the labor force. Given these adjustment costs, a minimum cost employment policy for a firm may be one in which the firm employs more labor during periods of declining output than is required to produce the desired level of output. If this hypothesis is true, then during periods of declining output the actual labor used in the production process may in fact be declining in proportion or more than in proportion to changes in the output even though measured employment does not. At the same time, once output begins to recover, the increase in measured employment will be less than proportionate to the increase in output because the firm can now use in its production the excess labor that was accumulated during the previous periods of declining output. The fact that measured employment may not be a very good measure of the true production function input could very well account for the contradictory results obtained by the earlier productivity and short-run employment studies.

The purpose of this paper is to construct a model of short-run employment behavior in durable goods industries based on an assumption of labor hoarding. The paper is composed of four sections. In Section I the labor-hoarding model is developed assuming perfect foresight, and the necessary and sufficient conditions for a minimum cost employment policy are established. Using the cost minimization conditions, a programming algorithm is derived which permits us to obtain estimates of the maximum length of time it would be profitable for firms to hoard excess labor, referred to as the labor-hoarding lag. Two models are constructed. The first model

* Assistant professor, University of Western Ontario.
I am indebted to Robert E. Hall for many helpful comments and criticisms, and to Stephen B. Browne for invaluable research assistance.

assumes there is no production lag in durable goods industries, while the second model allows for a production lag in the determination of required employment. In Section II the data are described. Section III presents the estimates of the labor-hoarding lags. Several hypotheses are also tested. Section IV contains the summary and conclusions.

I. The Models

We assume that the objective of the firm is to produce output $Q_1, Q_2, \ldots, Q_T$ over T periods at least cost. We also assume that the production function of the firm is such that a fixed number of workers is required per unit of output.[1] That is:

$$L_t \geqq \lambda Q_t \tag{1}$$

where L_t is the number of workers employed by the firm during period t, Q_t is the firm's output, and λ the production coefficient or labor-output ratio.[2]

We postulate that there are three costs associated with using labor: a wage, w, paid each period; a hiring cost, h, paid each time the labor force is enlarged; and a layoff cost, f, paid each time the labor force is reduced.[3] The firm's total labor cost over T periods is thus:[4]

$$\begin{aligned} C = {} & w(L_1 + L_2 +, \ldots, + L_T) \\ & + h([L_1 - L_0]^+ \\ & + [L_2 - L_1]^+ +, \ldots, \\ & + [L_T - L_{T-1}]^+) + f([L_0 - L_1]^+ \\ & + [L_1 - L_2]^+ +, \ldots, \\ & + [L_{T-1} - L_T]^+) \end{aligned} \tag{2}$$

where

$$\begin{aligned} [X]^+ &= 0 \qquad \text{if } X < 0 \\ &= X \qquad \text{if } X \geqq 0 \end{aligned}$$

Given (1) and (2), the necessary and

[1] The use of a fixed coefficient production function implies that capital and labor are not substitutable and therefore that variation in relative factor prices are not important in determining factor intensities. The large body of empirical work on production functions indicates, however, that manufacturing industries are characterized by factor substitutability. This would appear to indicate that the assumption of constant factor proportions for determining required employment is inappropriate. However, this is not necessarily so. It is not a contradiction to believe the empirical evidence which indicates that the true production technology is characterized by factor substitutability and at the same time still believe that, for the purpose of deducing factor demand from output demand, a fixed coefficient technology provides a satisfactory approximation to the true technology. Factor substitutability may not be important for the purpose of deducing factor demand if relative factor prices do not change very much. There are, of course, some practical reasons for using the fixed coefficient approximation. First, because of the estimation technique used, it is important to keep the number of structural parameters to be estimated to a minimum. Second, data for three-digit SIC industries are used in this paper, and there are, to date, no capital data available at this level of disaggregation.

[2] In constructing both labor-hoarding models in this paper, we have not allowed for substitution between the stock variable, workers, and the flow variable, the number of hours worked per worker. In our models, an increase or decrease in shipments and new orders necessitates an increase or decrease in required workers. This excludes the possibility of simply changing the utilization of the existing stock of workers (i.e., the number of hours worked per worker) in response to a change in shipments or new orders.

[3] Hiring costs are usually defined as including recruiting costs, orientation costs, recalling costs, training costs, and so forth. Layoff costs will include severance pay, that part of training costs that will not be recouped if the worker is laid off, and also unobservable costs such as bad worker morale, bad community relations, and union conflicts.

[4] We could have included a discount rate in our cost function. Assuming a constant discount rate, r, total discounted labor cost is:

$$\begin{aligned} C = {} & \sum_{t=1}^{T} \frac{wL_t}{(1+r)^t} + \sum_{t=1}^{T-1} \frac{h[L_t - L_{t-1}]^+}{(1+r)^t} \\ & + \sum_{t=1}^{T} \frac{f[L_{t-1} - L_t]^+}{(1+r)^t} \end{aligned}$$

The inclusion of the discount rate does not affect the derivation of the labor-hoarding lag t*. We would now have for any t that,

$$\frac{(h+f)/(1+r)^t}{w/(1+r)^t} = \frac{h+f}{w} = t^*$$

That is to say, when considering the ratio of two nominal values discounted at the same rate, we need consider only the ratio of the two nominal values.

sufficient conditions for a cost minimization employment policy can be derived as follows. In general, a firm will hold no more labor than is required to produce current output, except that workers will not be laid off for a temporary drop in output. The maximum length of time that is profitable for a firm to keep a worker on the payroll without having him work is given by the ratio of the layoff plus hiring cost to the wage cost. That is:

$$t^* = \frac{h + f}{w} \tag{3}$$

In other words, a minimum cost employment policy for a firm is one in which the firm will never have more than t* successive periods of positive excess labor. At the same time, in order for employment costs to be minimized, a layoff by a firm will never be followed by a hire within t* periods, since in that case it would be cheaper for the firm to eliminate both the layoff and the hire and carry excess labor. These two conditions are necessary and sufficient for a cost minimization employment policy. The interpretation of t* as the maximum length of time for which it would be profitable for a firm to hoard idle workers implies that the firm has perfect knowledge of its future output and consequently, from the production function, the path of required employment. This assumption of perfect foresight is implied in the specification of the cost function which includes future values of labor inputs. Without such knowledge, the firm would be unable to know exactly when it should begin hoarding labor or how much labor it should hoard. For the durable goods industries used in this study, the assumption of perfect foresight can be defended on two grounds. First, these durable goods industries produce primarily to order, and the lag between orders and production is sufficiently long to enable the firm to use its orders data in planning layoffs and hiring. Second, on the basis of past experience, firms have sufficiently good knowledge of the seasonal behavior of orders and shipments to permit accurate employment planning. If firms are fairly accurate in their forecasting, then the perfect foresight assumption might be a better approximation of how expectations are formed than hypotheses of imperfect expectations.[5]

An easy algorithm for estimating the labor-hoarding lag, t*, and the production coefficient, λ, is suggested by the two cost minimization conditions. We begin by taking as our initial guess of the firm's employment policy, the policy of holding no excess labor. In this case, we have from the production function that:

$$L_t = \lambda Q_t \tag{4}$$

Given data on output, and assuming a value for λ, we calculate the employment series that would occur if no excess labor were held. We then assume a value for t* and consider the earliest layoff (decline in employment) in the no excess labor employment series. If a hire (increase in employment) occurs within t* periods after the layoff, then the layoff and the hire are cancelled and the labor force increased for every intermediate period. This process is repeated over the entire sample until no hire occurs within t* periods of any layoff. At every iteration, the second cost minimization condition is satisfied so that the process will reach a minimum cost policy in a finite number of iterations. We select the maximum likelihood estimates of t* and λ from the set of all possible

[5] One method by which uncertainty could be introduced into the model would be to specify the production function in terms of expected output and expected employment. Alternative expectational hypotheses could be derived in terms of actual output. Such a procedure would certainly be necessary for industries which produce to stock but not necessarily for those which produce to order, which is the case for the industries used in this paper.

combinations by comparing the sum of squared residuals between the observed employment series and the estimated employment series based on our cost minimization conditions.

A brief examination of the employment-output data indicates that for most of the industries there has been a downward trend in the labor-output ratio. Having assumed away capital-labor substitution, there are two other possible explanations for this behavior. First, it is quite possible that the production function is one of increasing returns to scale rather than constant returns to scale. That is:

$$L_t \geqq \lambda Q_t^{\eta}, \qquad \eta < 1 \tag{5}$$

A second possible explanation is that there has been labor-augmenting technological change during the sample period. Allowing for this possibility, the technology approximation becomes:

$$L_t \geqq \lambda(t) Q_t \tag{6}$$

In this paper we adopt the second possibility. We assume that technological change takes place smoothly at a constant rate m, so that:

$$\lambda(t) = \lambda_0 e^{-mt} \tag{7}$$

Equations (6) and (7) simply say that a given level of output can now be obtained from a given capital input and a labor input, L_t, measured in men, that decreases over time. Within this model, a decreasing number of men represents an increasing number of labor efficiency units so that a fixed coefficient production relation is maintained between capital and labor efficiency units.

The labor-hoarding lag, t*, will be estimated using data for five three-digit SIC durable goods industries. By using three-digit industry data, it is possible to avoid some of the problems that might be caused by aggregation. It is very likely that many of the diverse relationships that exist between industries would be concealed if aggregate data were used. The use of three-digit industry data, however, does have some costs associated with it. Most important inventories by stage of fabrication are not available for three-digit industries. Consequently, it is necessary to use shipments as a measure of output. Since most of the durable goods industries used in this paper produce primarily to order, the assumption of no finished goods inventories is not unreasonable.[6] It is difficult to determine, however, what change in our output measure would occur if we were able to include the change in work-in-process inventories. It is quite possible that a decline in shipments could be completely offset by a build-up in work-in-process inventories so that output, as measured by shipments plus the change in work-in-process inventories, would show no decline, whereas output, as measured by shipments, would. This possibility would be particularly true for industries with fairly long production lags. If production of a good takes several periods, then the amount of labor required in any period will depend not only on the goods which reach completion and are shipped, but also on those goods still in the production process at different stages of completion (i.e., work-in-process inventories). To assume that required labor services depend only on shipments in the current period implies that there is no production lag.

There is sufficient evidence to indicate that in the durable goods industries production does take time. Consequently, in order to compensate for the possible inadequacy of shipments as a measure of output in this type of industry, we will develop a second labor-hoarding model based on the assumption of a production

[6] The possible exceptions are the blast furnaces and steel mills industry and the electrical distribution equipment industry.

lag. The second model is developed for an industry which produces $q+1$ different goods. Each good is distinguished by its production time, defined as the time between initiation of production and the completion of the final product.

We define the following variables:

$X_{t,j}$ = The value of production initiations of good j in period t

$L_{t,j}$ = The number of workers required during period t for the production of good j

As before, we assume that a fixed coefficient technology provides a satisfactory approximation to the true industry production function. More specifically, we assume that for good j a fixed number of workers is required per unit of production initiations of good j.

Consider now the demand for workers in period t to be used in the production of good j. In period t there will be some of good j reaching the final stage of production. These will be goods whose production was initiated in period $t-j$ (i.e., $X_{t-j,j}$). These goods will require some labor services. At the same time, there will be some of good j which will reach the final stage of production in period $t+1$. These goods whose production was initiated in period $t-j+1$ will also require some labor input. Thus, for each production of good j initiated in the current period and over the past j periods, there will be some labor input required in period t. Employment required for the production of good j in period t can then be written as a distributed lag function of production initiations of good j:

$$L_{t,j} = \sum_{\tau=0}^{j} \beta_\tau^{L,j} X_{t-\tau,j} \tag{8}$$

For $j=0$, the lag function would collapse to include only $X_{t,0}$:

$$L_{t,0} = \beta_0^{L,0} X_{t,0} \tag{9}$$

For $j=1$, we would have the relation:

$$L_{t,1} = \beta_0^{L,1} X_{t,1} + \beta_1^{L,1} X_{t-1,1} \tag{10}$$

Since there are $q+1$ goods there would be $q+1$ distributed lag labor requirement functions. Using a lag operator Z, equation (8) can be rewritten as:

$$L_{t,j} = \sum_{\tau=0}^{j} \beta_\tau^{L,j} Z^\tau X_{t,j} \tag{11}$$

Letting $\beta^{L,j}(Z) = \sum_{\tau=0}^{j} \beta_\tau^{L,j} Z^\tau$ we have the final labor requirement function for good j:

$$L_{t,j} = \beta^{L,j}(Z) X_{t,j} \tag{12}$$

Total labor required by the industry in period t is the sum of the labor requirements for each of the $q+1$ goods:

$$L_t = \sum_{j=0}^{q} \beta^{L,j}(Z) X_{t,j} \tag{13}$$

We now make the aggregation assumption that the mix of production initiations (i.e., $X_{t,0}, X_{t,1}, \ldots, X_{t,q}$) is constant for all t. That is:

$$X_{t,j} = \beta_j^p X_t \qquad (j = 0, 1, \ldots, q) \tag{14}$$

where X_t is the value of total production initiations in period t and the β_j^p measure the relative importance of the production initiations of each good. By definition:

$$\sum_{j=0}^{q} \beta_j^p = 1 \tag{15}$$

Substituting the aggregation assumption into (13) gives:

$$L_t = \sum_{j=0}^{q} \beta^{L,j}(Z) \beta_j^p X_t \tag{16}$$

Letting $\beta^L(Z) = \sum_{j=0}^{q} \beta^{L,j}(Z)\beta_j^p$ we have the final distributed lag relation between total labor requirements and the value of total production initiations:

$$L_t = \beta^L(Z) X_t \tag{17}$$

In the labor requirement equation, (17), the key variable (production initiations) is not observable. To complete the model for empirical testing we must relate X_t to some observable variable. In this paper, X_t is equated to the value of new orders received during period t. The labor requirement function then becomes:

$$L_t = \beta^L(Z)N_t \tag{18}$$

The lag function, $\beta^L(Z)$, is assumed to be stationary. The choice of new orders as a measure of production initiations in any period could result in situations, however, in which this assumption is not valid. If an industry is operating at less than full capacity, then any new orders received during a period can be initiated into production almost immediately. On the other hand, when capacity is being fully utilized, new orders will not be initiated immediately but instead will be placed in a queue to await production sometime in the future. Consequently, in addition to the work-time lag, there is now a wait-time lag. The latter lag component will vary over time depending on the degree of capacity utilization. In order for the stationarity assumption to be valid, we must assume that there is always sufficient excess capacity in the industries so that the wait-time lag is zero.[7] There are other reasons why the lag function may not be stationary. Even if there is excess capacity so that there is no queue, the work-time lag may change. First, there may be technological improvements resulting in a shortening of the work-time required for production. Secondly, there may be a cyclical response in work time. As demand conditions improve, firms may be able to shorten their work time by increasing their work speed. However, if demand continues to rise at a fairly rapid rate, there may be a relative deterioration in the earlier improvement as less efficient capital and labor must be employed. Finally, there may be a change in the relative importance of the various commodities in total demand.[8]

We can now complete the second labor-hoarding model. We assume that the objective of the firm is to fill all new orders $N_1, N_2, \ldots, N_T$, over T periods at least cost. Based on the assumption of a production lag and a fixed coefficient technology, the number of workers employed in period t is given as:

$$L_t \geqq \beta^L(Z)N_t \tag{19}$$

The cost function, (2), used in the first model is also applicable in this model. Since only the labor requirement function has been changed, the necessary and sufficient conditions for a cost minimization employment policy are essentially the same as those in the first model. In general, a firm will hold no more labor than is required to complete the production of

[7] This problem has arisen in other empirical work dealing with the distributed lag relationship between shipments and new orders. A number of investigators in this area attempted to allow for nonstationarity by using an estimation method known as variable weighted distributed lags (see Joel Popkin and Peter Tinsley). This approach to the problem is somewhat *ad hoc* in that models are not usually based on a convincing underlying behavioral hypothesis. A second alternative is to allow for varying capacity utilization through the specification of production initiations. For example, we might let $X_t = \text{Min}\,(C_t, Q_t)$ where C_t is capacity and Q_t the queue of unfilled orders for which production has not been initiated at the beginning of period t. This model implies a switching between two functions; when there is excess capacity, production initiations are equal to Q_t; when there is a capacity constraint, production initiations are then equal to C_t.

[8] It is easy to demonstrate that the labor requirement function in (4) is a constrained form of the more general case given in (18). If we adopt the assumption that there is no production lag there will be only one good produced in the industry (i.e., good 0). This implies that $X_{t,0} = X_t$. Equating production initiations to new orders gives $X_t = N_t$. Since we are also assuming no wait-time lag, every new order in period t becomes a shipment in period t: $N_t = S_t$. The labor requirement function is then simply $L_t = \beta_0^{L,0} S_t$ which is identical to equation (4).

those goods in the final stages of processing:

$$N_{t,0},\ N_{t-1,1},\ \ldots,\ N_{t-q,q} \tag{20}$$

and also labor required to maintain work on goods which will reach completion in periods t+1, t+2, . . . , t+q:

$$\begin{aligned} &N_{t,1} \\ &N_{t,2},\ N_{t-1,2} \\ &\vdots \\ &N_{t,q},\ N_{t-1,q},\ \ldots,\ N_{t-q+1,q} \end{aligned} \tag{21}$$

On the other hand, workers will not be laid off for a temporary drop in new orders. The maximum length of time that is profitable for a firm to keep a worker on the payroll without having him work is given by the ratio of the layoff plus hiring cost to the wage cost:

$$t^* = \frac{h+f}{w} \tag{22}$$

The necessary and sufficient conditions for a minimum cost employment policy are the same as in the first model. A firm will never have more than t* successive periods of excess labor. Moreover a layoff by a firm will never be followed by a hire within t* periods, since it would then be cheaper to eliminate both the layoff and the hire and carry excess labor.

To estimate the labor-hoarding lag, t*, we employ the algorithm described above. Three different specifications of $\beta^L(Z)$ will be used. To begin with, we adopt the simple specification that the lag coefficients correspond to a rectangular lag function. This can be expressed as:

$$\begin{aligned} \beta_\tau &= \alpha \qquad \text{if } 0 \leqq \tau \leqq q \\ &= 0 \qquad \text{otherwise} \end{aligned} \tag{23}$$

The labor requirement function then becomes:

$$L_t = \alpha \sum_{\tau=0}^{q} N_{t-\tau} \tag{24}$$

leaving only three parameters to be estimated: α, q, and t*.

Assuming values for α and q, we obtain an estimate of the no excess labor series from (24). With this series, we are then able to estimate the labor-hoarding lag, t*, using the cost minimization algorithm. Initial estimates of α, for different values of q, can be obtained by dividing the observed employment series by $\sum_{\tau=0}^{q} N_{t-\tau}$. Once having obtained the best estimates of α and t*, for each q, we select as our best estimates of all three parameters that equation which has the smallest variance corrected for degrees of freedom.

The second lag function to be used in this study is the inverted V or symmetric triangle. Using this specification, the lag coefficients are given by:

$$\begin{aligned} \beta_\tau &= \alpha_1\tau & 0 \leqq \tau \leqq \alpha_2 \\ &= \alpha_1\alpha_2 - \alpha_1(\tau - \alpha_2) & \alpha_2 < \tau \leq 2\alpha_2 \end{aligned} \tag{25}$$

where α_1 is the slope of the triangle and $2\alpha_2$ its length. Substituting this specification into the labor requirement function gives:

$$\begin{aligned} L_t = \alpha_1 &\sum_{\tau=0}^{\alpha_2} \tau N_{t-\tau} \\ &+ \sum_{\tau=\alpha_2+1}^{2\alpha_2} [\alpha_1\alpha_2 - \alpha_1(\tau - \alpha_2)] N_{t-\tau} \end{aligned} \tag{26}$$

Since the triangle is symmetric, this is equivalent to:

$$\begin{aligned} L_t = \alpha_1 &\sum_{\tau=0}^{\alpha_2-1} \tau (N_{t-\tau} + N_{t-2\alpha_2+\tau}) \\ &+ \alpha_1\alpha_2 N_{t-\alpha_2} \end{aligned} \tag{27}$$

In the case of the triangle there are three parameters to estimate: the slope α_1, the mid-point of the triangle α_2, and t*.

The estimation of the model with a triangular lag function is similar to that with the rectangle. Assuming values for α_2 and α_1, we obtain an estimate of the no excess labor series from (27). The labor-

hoarding lag, t*, is then estimated as before. We select as our best estimates of α_1, α_2, and t* that equation with the smallest variance corrected for degrees of freedom. Initial estimates of the slope, α_1, for different values of α_2, can be obtained by dividing observed employment series by:

$$\sum_{\tau=0}^{\alpha_2-1} \tau(N_{t-\tau} + N_{t-2\alpha_2+\tau}) + \alpha_2 N_{t-\alpha_2} \tag{28}$$

The third distributed lag function to be used in our labor requirement equation is the common geometric function. Under this specification, the coefficients are given by:

$$\beta_\tau = \alpha_1 \alpha_2^\tau \qquad 0 \leq \alpha_2 < 1 \tag{29}$$

The labor requirement function becomes:

$$L_t = \alpha_1 \sum_{\tau=0}^{\infty} \alpha_2^\tau N_{t-\tau} \tag{30}$$

Using a Koyck transformation, we obtain the labor requirement equation to be used in estimation:

$$L_t = \alpha_1 N_t + \alpha_2 L_{t-1} \tag{31}$$

With the geometric lag function, there are three parameters to be estimated; α_1, α_2, and t*. Initial estimates of α_1 and α_2 can be obtained in the following manner. First, we know that α_2 must be less than one and equal to or greater than zero. Assuming a value for α_2 within these bounds, we obtain an estimate of the range of α_1 by dividing observed employment by:

$$\sum_{\tau=0}^{K} \alpha_2^\tau N_{t-\tau} \tag{32}$$

where K is a constant inserted to make the sum finite. The larger K is, the better (32) approximates the infinite sum in (30).

There is, however, one additional estimation problem. It is necessary in the case of the geometric lag function to have an estimate of the number of nonidle workers held in period zero (i.e., L_0). In other words, it is necessary to consider different initial values for for L_0 and for each value obtain the best estimates of α_1, α_2, and t*. We select as our best estimates of L_0, α_1, α_2, and t*, that equation which has the smallest variance corrected for degrees of freedom.

So far in our discussion of this model, we have not considered technological change. We now assume that technological change takes place smoothly at a constant rate, m, so that employment is now given by:

$$L_t \geq \lambda_0 e^{-mt} \beta^L(Z) N_t \tag{33}$$

This results in one additional parameter to be estimated for each of the three distributed lag functions.

With technological change and a rectangular lag function, the labor requirement equation becomes:

$$L_t = \alpha e^{-mt} \sum_{\tau=0}^{q} N_{t-\tau} \tag{34}$$

In the case of the triangular lag function, using (27), we have:

$$L_t = \alpha_1 e^{-mt} \left[\sum_{\tau=0}^{\alpha_2-1} \tau(N_{t-\tau} + N_{t-2\alpha_2+\tau}) + \alpha_2 N_{t-\alpha_2} \right] \tag{35}$$

For the geometric lag function, the inclusion of technological change gives:

$$L_t = \alpha_1 e^{-mt} \sum_{\tau=0}^{\infty} \alpha_2^\tau N_{t-\tau} \tag{36}$$

Employing a Koyck transformation we obtain:

$$L_t = \alpha_1 e^{-mt} N_t + \alpha_2 L_{t-1} \tag{37}$$

The introduction of technological change in this manner allows the lag func-

tions in the labor requirement equation to be semi-nonstationary. For example, in the case of the rectangle, the inclusion of technological change implies that the height of the rectangle is decreasing over time, whereas in the case of the triangle, it implies that the slope is decreasing over time. In both cases, however, the length of the production lag remains constant. For the geometric lag function, the inclusion of technological change implies that the height of the function is declining.

In Section III estimates of the labor-hoarding lag are presented based on equations (6), (34), (35), and (37).

II. Data

The labor-hoarding lag, t^*, is estimated for the following three-digit SIC durable goods industries:

331—Blast furnaces and steel mills[9]
351—Engines and turbines
353—Construction, mining, and material handling equipment
354—Metal working machinery
361—Electrical distribution equipment

The shipments and new orders data are monthly unseasonally adjusted data obtained from the monthly survey of manufacturers published by the U. S. Department of Commerce. The data were not seasonally adjusted since the purpose of this paper is to examine the short-run behavior of employment. The sample used in this study covers the period January 1960 to January 1969. This includes one recession, 1960–61; the moderate expansion beginning late in 1961 and culminating in the more rapid expansion of 1965 and early 1966; the mini-recession of 1967 followed by a period of moderate growth until the beginning of 1969.

To obtain constant dollar shipments and new orders series, wholesale price indices published by the Bureau of Labor Statistics (*BLS*) were used. Unfortunately, the industry classifications used by *BLS* are not the same as the SIC classifications used by the Department of Commerce for its shipments data. It was not until 1967 that the Bureau of Labor Statistics began to publish wholesale price indices on an SIC basis. As a result, it was necessary to combine *BLS* price indices which would represent as much as possible the SIC industry.

Employment data are available for SIC industry classifications. There were some differences, however, between the industry classifications used by the Department of Labor and the classifications used by the Department of Commerce. The differences applied only to industry 354, metal working machinery, and industry 361, electrical distribution equipment. Fortunately, however, the employment data are available for four-digit industry classifications so that it was possible to derive employment series for industries 354 and 361 which are comparable to the industry classifications on which the shipments and new orders data are based. Both total employment and employment of production workers are available for our sample period. An estimate of employment of nonproduction workers is obtained as the difference between total employment and employment of production workers.

III. Empirical Results

Table 1 contains the best estimates of the labor-hoarding lag, t^*, for the four models estimated.[10] For the model with-

[9] The name of this industry classification is misleading. Unlike the other industries, the name of industry 331 does not refer to the output but to the means used to produce output. The output of this industry consists of ferrous and nonferrous alloy and finished and semi-finished steel products.

[10] The ranges over which the parameters were allowed to vary were large enough to insure that the best estimates were obtained. The ranges differed, of course,

[over]

TABLE 1—ESTIMATES OF THE LABOR-HOARDING LAG, t*, IN MONTHS, AND TESTS OF THE NULL HYPOTHESIS THAT t*=0

				With Production Lag								
	Without Production Lag			Rectangle			Triangle			Geometric		
Industry	t*	s	Z[b]	t*	s	Z[b]	t*	s	Z[b]	t*	s	Z[b]
				Total Employment								
331	14	948.87	1,843	3	1,855.22	12	7	2,755.99[a]	43	2	680.13	0
351	5	34.96	1,166	15	12.18	1,822	17	22.65	768	14	15.94	1,931
353	9	32.19	2,675	8	32.43	47	17	65.94	843	15	36.44	1,303
354	6	21.56	14,148	12	129.08	13	3	143.51[a]	0	12	8.76	20
361	7	73.11	477	9	11.03	341	17	4.92	297	7	19.97	1,100
				Production Workers								
331	14	897.56	1,081	6	1,038.10	81	7	1,844.47	39	6	631.75	8
351	5	22.39	707	16	11.23	608	16	16.34	243	14	11.31	1,341
353	9	17.54	1,909	17	9.89	18,741	7	12.77	1,739	2	13.99	127
354	5	12.21	1,061	14	59.09	16	4	61.35[a]	0	19	6.80	111
361	5	29.43	481	12	8.10	68	14	3.11	259	10	9.23	1,031
				Nonproduction Workers								
331	14	57.30	1,397	9	89.67	33	16	122.87[a]	82	22	27.92	560
351	9	1.71	4,539	15	.58	2,241	19	.92	4,352	13	.64	500
353	11	7.20	1,991	17	19.54	195	16	20.91	543	16	3.85	417
354	8	3.40	1,192	14	22.27	25	1	21.43[a]	22	22	1.86	6
361	7	10.45	405	15	.81	363	20	.79	903	9	1.54	1,813

Note: s = Variance adjusted for degrees of freedom computed from the sum of squared residuals between the observed employment series and the estimated employment series based on the cost minimization conditions.

[a] These estimates are not based on complete convergence of the variance (see fn. 10).

[b] The critical value of $\chi^2(1)$ at the .05 level is 3.84. Equation (38) in the text defines the test-statistic Z.

out a production lag, estimates of the labor-hoarding lag ranged from 5 to 14 months in the case of both total employment and employment of production workers. For employment of nonproduction workers, the estimated labor-hoarding lags ranged from 7 to 14 months. For the rectangular lag function, estimates of the labor-hoarding lag ranged from 3 to 15 months for total employment, 6 to 17 months for production workers, and 9 to 17 months for nonproduction workers. For the triangle, the corresponding ranges were 3 to 17 months for total employment, 4 to 16 months for production workers, and 1 to 20 months for nonproduction workers. Finally, for the geometric lag function, estimates of the labor-hoarding lag ranged from 2 to 15 months for total employment, 2 to 19 months for employment of production workers, and 9 to 22 months for employment of nonproduction workers.[11] Although there is not that much

from function to function since in some cases convergence was fast whereas in other cases it was slow. For the labor-hoarding lag, t*, the range was always 0 to 30 months. For the triangular lag function, five of the equations are not based on complete convergence of the variance about the estimated regression. In each of these five employment series, it was found that for large values of α_2 the estimated minimum cost employment series underestimated observed employment at the beginning of the samples for a number of months. Consequently, as α_2 was increased and the sample size reduced, the sums of squared residuals continued to decline. The estimates of the labor-hoarding lag, however, did not appear to be very sensitive in these series to increases in α_2 beyond 9 or 10 months. It was decided, therefore, to stop estimation at this point with the result that although the estimates of t* are fairly close to their maximum likelihood estimates, the estimates of the production lag probably contain some downward bias.

[11] It was found that the estimates of the parameters were not very sensitive to the choice of L_0 in the vicinity

TABLE 2—PARAMETER ESTIMATES AND TESTS OF THE NULL HYPOTHESIS THAT THE PRODUCTION LAG IS NOT IMPORTANT IN EXPLAINING THE SHORT-RUN DEMAND FOR EMPLOYMENT

	No Production Lag		Rectangle			Triangle			Geometric			Test that Production Lag is Zero	
Industry	λ_0	m	α	q	m	α_1	α_2	m	α_1	α_2	m	Best Function	Z[a]
						Total Employment							
331	.34	.003	.040	10	.004	.0666	8	.004	.045	.90	.004	*G*	42
351	.48	.005	.161	2	.006	.055	3	.006	.31	.31	.005	*R*	206
353	.53	.002	.039	14	.002	.057	3	.001	.15	.70	.0004	*G*	0
354	.76	.004	.051	14	.004	.0091	9	.003	.030	.96	.002	*G*	152
361	.32	.001	.029	12	.004	.0076	7	.004	.28	.20	.004	*T*	1,437
						Production Workers							
331	.28	.003	.029	11	.004	.0054	8	.004	.037	.89	.003	*G*	44
351	.31	.004	.079	3	.005	.014	5	.006	.19	.30	.004	*R*	119
353	.34	.001	.022	12	.002	.0049	9	.002	.038	.90	.001	*R*	79
354	.54	.004	.039	12	.003	.0063	9	.003	.030	.94	.002	*G*	88
361	.22	.001	.020	11	.003	.0067	6	.003	.16	.29	.003	*T*	97
						Nonproduction Workers							
331	.06	.002	.0058	13	.004	.0012	8	.004	.0075	.90	.003	*G*	106
351	.16	.007	.032	5	.009	.019	3	.008	.11	.29	.008	*R*	248
353	.17	.002	.019	9	.002	.043	2	.002	.013	.93	.002	*G*	76
354	.22	.004	.019	11	.004	.0025	9	.003	.009	.96	.003	*G*	88
361	.11	.002	.0090	13	.005	.0050	5	.005	.094	.20	.005	*T*	1,283

Note: R = Rectangular lag function; T = Triangular lag function; G = Geometric lag function.
[a] The critical value of $\chi^2(1)$ at the .05 significance level is 3.84.

difference in the range of estimates for the three employment classifications across functions, the estimates of the labor-hoarding lags for individual industries do appear to be quite sensitive to the specification of the distributed lag function. For example, in the case of employment of nonproduction workers in the metal working machinery industry, the estimates of t* ranged from 1 month in the case of the triangle to 22 months in the case of the geometric lag function. For blast furnaces and steel mills the estimates of the labor-hoarding lag for nonproduction workers ranged from 9 months in the case of the rectangle to 22 months in the case of the geometric lag function.

Table 2 contains the estimates of the additional parameters for each of the models. We can derive estimates of the production lag by adding 1 to the estimate of q in the case of the rectangle, and by multiplying α_2 by 2 in the case of the triangle. For the rectangle, the production lag estimates obtained from the three employment series did not vary much for the individual industries. For the triangle, there was a larger variation in the estimates. For example, in the case of industry 353, the estimated production lag from the nonproduction worker series was 4 months, from production workers 18 months, and from total employment 6 months. Finally, our results indicate that the production lag estimates, like the estimates of the labor-hoarding lags, are

of the beginning value of the observed employment series. For this reason L_0 was fixed at a value very close to the beginning value of the observed employment series.

fairly sensitive to the specification of the lag function.

We now ask whether our estimates of t*, for both the model without a production lag and the model with a production lag, provide strong enough evidence for rejecting the hypothesis that there is no labor hoarding (i.e., t*=0). A test of this hypothesis can be made by constructing the statistic:

$$Z = \frac{S_0^2 - S_1^2}{s} \tag{38}$$

which is asymptotically $\chi^2(1)$. The term S_0^2 is the sum of squared residuals under the null hypothesis that $t^*=0$; S_1^2 is the unconstrained sum of squared residuals; and s is an estimate of the variance around the regression from the unconstrained equation. The estimates of Z are given in Table 1. For the model without a production lag, the values of Z provide strong evidence for rejecting the null hypothesis at the 5 percent significance level in every equation. For the model with a production lag, in only three of the forty-five equations estimated are the values of Z such that we cannot reject the hypothesis of no labor hoarding.

The second hypothesis that we can test is the hypothesis that the production lag is not important in explaining the short-run demand for employment in the durable goods industries. A test of this hypothesis can be made by using the test statistic (Z) which will be distributed asymptotically as $\chi^2(1)$.[12] The term S_0^2 will now be the sum of squared residuals under the null hypothesis that there is no production lag; S_1^2 will be the sum of squared residuals from the unconstrained model which includes a production lag; and s is the estimate of the variance around the regression from the unconstrained equation. For the model with a production lag, the equation with the smallest variance adjusted for degrees of freedom was used in the test. Looking at Table 1, we can see that the geometric lag function performed best for all three employment series in the case of industries 331 and 354. The rectangular and triangular lag functions performed best for all three employment series for industry 351 and industry 361, respectively. For industry 353, the geometric lag function provided the best estimates in the case of total employment and employment of nonproduction workers but the rectangular lag function performed best in the case of production workers. The estimated values of Z given in Table 2 provide strong evidence for rejecting the null hypothesis at the 5 percent significance level in fourteen of the fifteen equations.

Finally, since we have estimated the employment model for total employment and its two components, employment of production workers and employment of nonproduction workers, it is possible to test the hypothesis that production workers and nonproduction workers are the same. This test can be made by computing the statistic:

$$V = \frac{S_{TE}^2 - (S_{PW}^2 + S_{NPW}^2)}{s_{PW} + s_{NPW}} \tag{39}$$

which is distributed asymptotically as $\chi^2(3)$ for the model without a production lag and $\chi^2(4)$ for the model with a production lag.[13] The term S_{TE}^2 is the sum of squared residuals from the total employ-

[12] If the production lag is not important, then this is equivalent to saying that $q=0$ in the case of the rectangle; that $\alpha_1=0$ in the case of the triangle; and, that $\alpha_2=0$ in the case of the geometric function.

[13] For the model with a production lag, we are assuming that there are four parameters identical: α_1, q, t*, and m in the case of the rectangle; α_1, α_2, t*, and m in the case of the triangle; and α_1, α_2, t*, and m in the case of the geometric lag function. For the model without a production lag, three parameters are assumed equal: α, m, and t*.

ment equation, representing the null-hypothesis that production workers and nonproduction workers are the same; S^2_{PW} and S^2_{NPW} are the sum of squared residuals from the unconstrained equations for production workers and nonproduction workers; and s_{PW} and s_{NPW} are estimates of the variances around the regressions from the unconstrained equations. The estimates of V are given in Table 3.

TABLE 3—TEST OF THE NULL HYPOTHESIS THAT PRODUCTION WORKERS AND NONPRODUCTION WORKERS ARE THE SAME

	No Production Lag	Assuming a Production Lag		
		V		
Industry	V	Rectangle	Triangle	Geometric
331	0	58	35	3
351	45	4	27	37
353	26	6	73	100
354	39	42	65	3
361	86	17	18	72

The critical values of $\chi^2(3)$ and $\chi^2(4)$ at the .05 significance level are 7.81 and 9.49, respectively.

In the case of the model without a production lag, the values of V for four of the five industries provide strong evidence for rejecting the hypothesis at the 5 percent significance level. For the model with a production lag, the null hypothesis is rejected in eleven of the fifteen equations.

IV. Summary and Conclusions

The purpose of this paper has been to construct a model of short-run employment behavior in durable goods manufacturing industries based on a theory of labor hoarding. This theory operates on the assumption that there are important costs associated with changing the labor force. Given these labor force adjustment costs, an optimal or minimum cost employment policy for a firm may be one in which the firm employs more labor during periods of declining output than is required to produce the desired level of output. Using the necessary and sufficient conditions for cost minimization, a programming algorithm was constructed which permitted us to obtain estimates of the maximum length of time it would be profitable for a firm to hoard idle workers, referred to as the labor-hoarding lag. Two models were constructed; the first model assumed that there was no production lag in durable goods industries, and the second model allowed for a production lag in the determination of required employment.

The results from both models provide strong evidence in support of the hypothesis that durable goods industries do hoard labor. Estimates of the labor-hoarding lag ranged from a low of 5 months to a high of 14 months in the case of the first model, and from a low of 1 month to a high of 22 months in the case of the second model. In addition, the performance of the second model in explaining the short-run behavior of employment in these industries was definitely superior to that of the first model. It was found, however, that the estimates of the labor-hoarding lag and production lag were fairly sensitive to the specification of the lag function.

REFERENCES

A. Ando, E. C. Brown, J. Kareken, and R. Solow, "Lags in Fiscal and Monetary Policy," in E. C. Brown et al., *Stabilization Policies*, Englewood Cliffs 1963.

R. J. Ball and E. B. A. St. Cyr, "Short-run Employment Functions in British Manufacturing," *Rev. Econ. Stud.*, July 1966, *33*, 179–207.

C. Bischoff, "The Lag Between Orders and Production of Machinery and Equipment: A Re-examination of the Kareken-Solow Results," working pap. no. 138, CRMS-

IBER, Univ. California, Berkeley, July 1968.

F. Brechling, "The Relationship Between Output and Employment in British Manufacturing Industries," *Rev. Econ. Stud.*, July 1965, *32*, 187–216.

——— and Peter O'Brien, "Short-Run Employment Functions in Manufacturing Industries: An International Comparison," *Rev. Econ. Stud.*, Aug. 1967, *49*, 277–97.

R. C. Fair, *Short-run Demand for Workers and Hours*, Amsterdam 1969.

R. E. Hall, "Transform Methods for the Study of Distributed Lags," working pap. no. 132, CRMS-IBER, Univ. California, Berkeley, June 1968.

E. Kuh, "Income Distribution and Employment Over the Business Cycle," in J. S. Duesenberry et al., eds., *Brookings Quarterly Econometric Model of the United States Economy*, Chicago 1965.

———, "Profits, Profit Markups and Productivity," study pap. no. 15, *Joint Economic Committee, Study of Employment, Growth and Price Levels*, Jan. 1960.

M. I. Nadiri, "The Effects of Relative Prices and Capacity on the Demand for Labor in the U.S. Manufacturing Sector," *Rev. Econ. Stud.*, July 1968, *35*, 273–88.

——— and S. Rosen, "Interrelated Factor Demand Equations," *Amer. Econ. Rev.*, Sept. 1969, *59*, 457–71.

M. Nerlove, "Notes on the Production and Derived Demand Relations Included in Macro-Econometric Models," *Int. Econ. Rev.*, June 1967, *8*, 223–42.

W. Y. Oi, "Labor as a Quasi-Fixed Factor," *J. Polit. Econ.*, Dec. 1962, *70*, 538–55.

J. Popkin, "The Relationship Between New Orders and Shipments: An Analysis of Machinery and Equipment Industries," *Sur. Curr. Bus.*, *45*, 24–32.

D. J. Smyth and N. J. Ireland, "Short-Term Employment Functions in Australian Manufacturing," *Rev. Econ. Stud.*, Nov. 1967, *49*, 537–44.

R. Solow, "Presidential Address to Econometric Society," mimeo. 1964.

P. A. Tinsley, "An Application of Variable Weighted Distributed Lags," *J. Amer. Statist. Assn.*, Dec. 1967, *62*, 1277–89.

U.S. Bureau of Labor Statistics, *Employment and Earnings Statistics for the United States, 1909–1968*, Bull. No. 1312-6, Washington 1968.

———, *Relative Importance of Individual Commodities*, Washington 1968.

———, *Wholesale Index Rebasing Factors*, Washington 1962.

———, *Wholesale Prices and Price Indices*, Bull. 1376, 1382, 1411, 1513, Washington 1963–65.

U.S. Department of Commerce, Bureau of Census, *Manufacturers' Shipments, Inventories, and New Orders:* 1947–1963, Revised Series M3-1, Washington 1963.

———, *Manufacturers' Shipments, Inventories, and New Orders:* 1961–1968, Series M3-1.1, Washington 1968.

Problems of Economic Policy from the Viewpoint of Optimal Control

By GREGORY C. CHOW*

A core problem in macro-economic policy is the evaluation of the performance of the economy under different specifications concerning the policy instruments. There are a number of interesting problems related to this core. The first is the setting of objectives in terms of which the performance can be evaluated. Second, given our knowledge of the economy, do policies exist that will serve to achieve or closely achieve these objectives? Third, how are good policies to be found and what form should they take? Presumably, a policy may involve coordinated use of several instruments. Fourth, assuming that our competing objectives, such as full employment and price stability, cannot be met perfectly by the best of policies, by how much can we improve upon the attainment of one objective by sacrificing the attainment of another? This is the problem of measuring the tradeoff possibilities available to the policy maker. Fifth, how are different policy proposals to be ranked and compared? Sixth, how can historical policies be appraised? Seventh, if one had to choose between subsets of instruments, say between monetary and fiscal instruments, in what way could a choice be made? In other words, how can the effectiveness of a subset of instruments be measured, relative to that of another subset?

In this paper, I hope to demonstrate that these problems can, and indeed should, be studied by the tools of optimal control for stochastic systems in economics. The above problems are by no means exhaustive, but they will be sufficient for illustrative purposes. In fact, the list may already be too long to be thoroughly treated in one paper. Fortunately for my exposition, a recent article by Robert Holbrook on "Optimal Economic Policy and the Problem of Instrument Instability" has raised several issues that are related to the first four problems listed in the preceding paragraph. I will use the topic of instrument instability as a starting and focal point of my discourse on the first four problems, leaving the remaining problems to be discussed later in this paper. I will show that the questions raised, and only partially studied, by Holbrook, as well as the other questions concerning optimal economic policy, can be answered by the techniques of optimal control in a fairly general setting.

Instrument instability was said to exist by Holbrook if "attempts to offset completely the cumulative impact of past changes in the policy instrument may require ever greater changes in the future value of the instrument" (p. 57). An example is provided by the model

$$y_t = .4x_t + .6x_{t-1} + 10$$

where y_t is the endogenous variable to be controlled and x_t is the instrument. If the objective is to make $y_t = 10$, for many future periods, the solution is to set $.4x_t + .6x_{t-1} = 0$, or $x_t = -1.5x_{t-1}$, a situation characterized as one of instrument instability.

* Princeton University. I would like to acknowledge, with thanks, helpful comments from George Borts, and financial support from the National Science Foundation through Grant NSF GS 32003X.

It is not difficult to see that the questions of instrument instability are related to the first four problems listed above. To study these problems in a more general setting than Holbrook, let there be given a linear econometric model in its reduced form,

$$(1) \quad y_t = A_1 y_{t-1} + \ldots + A_m y_{t-m} + C_0 x_t + \ldots + C_n x_{t-n} + b_t + u_t$$

where y_t is a vector of endogenous variables, x_t is a vector of instruments, A_i and C_i are given constant matrices, b_t are given vectors capturing the combined effects of other exogenous variables which are not subject to control, and u_t is a random vector with mean 0, covariance matrix V, and is uncorrelated with $u_s(t \neq s)$. One may ask, under what conditions will instrument instability exist? This question requires a definition of "instability," which involves the setting of objectives for the performance of the economy. Given some definition of stability as an objective, it is concerned with the possible nonexistence of a policy that will serve to achieve the desired objective. Secondly, one may ask, if instrument instability is found to exist, to what extent and in what way can the policy maker trade off this instability by allowing for instability in the endogenous variable(s) to be controlled? This question involves finding a good solution to reduce instrument instability, and ascertaining the tradeoff between instabilities in the instrument(s) and in the endogenous variable(s). Third, if one instrument, say federal expenditures, is found to be unstable when it is regarded as the only instrument (holding as fixed another instrument such as money supply), to what extent will the availability of another instrument alleviate the problem of instrument instability? This obviously involves the tradeoff between instabilities in different instruments. This question, which was not raised by Holbrook because his framework only allowed him to study one instrument at a time, will simply be treated by our method as a special case of the second question.

Our framework will be more general than Holbrook's in (a) having possibly more than one endogenous variable to be controlled, (b) having possibly more than one instrument to be studied, (c) having possibly unequal numbers of endogenous variables and instruments, and (d) having possibly random disturbances in the system. Furthermore, our solution can be generalized to the case of (e) treating the matrices A_i and C_i as random matrices when the policy maker is uncertain about their values, and (f) of non-linear dynamic models. This paper will deal only with generalizations (a), (b), (c), and (d), leaving (e) and (f) to appropriate references. All the questions raised so far can be analyzed fairly easily once a framework of optimal control is set forth. This is the immediate task of Section I. In Section II, the results of this framework will be used to study problems associated with instrument instability. In Section III, modifications due to the stochastic nature of the problem will be set forth. Section IV deals with the other problems of economic policy mentioned above.

I. A Framework of Optimal Control

Let us begin by rewriting the dynamic system (1) as a first-order system (2) in which the "current endogenous variables" will incorporate the instruments. We shall redesignate system (2) simply as equation (3).

$$(3) \quad y_t = A y_{t-1} + C x_t + b_t + u_t$$

Note that the newly defined y_t includes current and (possibly) lagged endogenous variables as well as current and (possibly) lagged instruments, whereas x_t remains the same as before. This rewriting serves

$$(2)\quad \begin{bmatrix} y_t \\ \cdot \\ \cdot \\ y_{t-m+1} \\ x_t \\ \cdot \\ \cdot \\ x_{t-n+1} \end{bmatrix} = \begin{bmatrix} A_1 \dots A_m & C_1 \dots C_n \\ & \\ \dots I \quad 0 & 0 \dots 0 \\ 0 \dots 0 & 0 \dots 0 \\ & \\ 0 \dots 0 & \dots I \quad 0 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ \cdot \\ \cdot \\ y_{t-m} \\ x_{t-1} \\ \cdot \\ \cdot \\ x_{t-n} \end{bmatrix} + \begin{bmatrix} C_0 \\ \cdot \\ \cdot \\ 0 \\ I \\ \cdot \\ \cdot \\ 0 \end{bmatrix} x_t + \begin{bmatrix} b_t \\ \cdot \\ \cdot \\ 0 \\ 0 \\ \cdot \\ \cdot \\ 0 \end{bmatrix} + \begin{bmatrix} u_t \\ \cdot \\ \cdot \\ 0 \\ 0 \\ \cdot \\ \cdot \\ 0 \end{bmatrix}$$

two main purposes. First, by incorporating the instruments in y_t, we can evaluate the performance of the economic system by the behavior of y_t alone even when the behavior of some instruments, such as government expenditures, should be included. This will simplify the writing of the objective function. Second, as is well known, the reduction of a higher order system into first order also facilitates exposition.

The performance of the system for a finite planning horizon from period 1 to period T will be judged by the deviations of y_t, as defined in (3), from the vectors a_t which are targets to be specified by the policy maker ($t=1, \dots, T$). Specifically, welfare cost will be measured by

$$(4)\quad W = E_0 \sum_{t=1}^{T} (y_t - a_t)' K_t (y_t - a_t)$$

where E_0 denotes expectation conditional on the initial condition y_0, again in the notation of (3), the prime denotes transpose, and K_t are known, symmetric (usually diagonal), positive semidefinite matrices, with zero elements as a rule corresponding to lagged (endogenous and control) variables. The problem is to minimize expected welfare cost (4) by setting the time path of the instrument $x_t(t=1, \dots, T)$, given the model (3).

The formulation of this problem is dynamic. The performance criterion evaluates y_t for many periods. Since the decision x_1 for the first period will affect the outcomes y_t for later periods through the system (3), thus restricting the possibilities open in later periods, one should take into account the multiperiod character of this problem in choosing (and hence also in appraising) the values of the instruments at each period. The problem is also stochastic. One cannot control the future outcomes exactly, but only the probability distributions of the outcomes and the expectation of certain (welfare) functions of these outcomes. It is concerned with the evaluation of stochastic time-series y_t under different regimes. The stochastic nature has implications on comparing different policy proposals and on judging historical policies.

Because the model is linear, its parameters are assumed to be known, and because the welfare cost is quadratic, the dynamic stochastic problem here formulated is one of the simplest and most basic problems in stochastic control theory. While other elementary methods of solving this problem can be found in Chow (1972a, b), I will provide here a solution using the method of dynamic programming of Richard Bellman. The method begins by solving the problem for the last period T, given the initial condition y_{T-1}. The welfare cost is

$$(5)\quad \begin{aligned} W_T &= E_{T-1}(y_T - a_T)' K_T (y_T - a_T) \\ &= E_{T-1}[y_T' H_T y_T - 2y_T' h_T + c_T] \end{aligned}$$

where we have set $H_T = K_T$ for ease of generalization to the multiperiod problem later on, and have set $h_T = k_T = K_T a_T$, and $c_T = a_T' K_T a_T$. Using the model (3) for y_T

and taking expectations, we minimize

$$(6)\quad W_T = (Ay_{T-1} + Cx_T + b_T)'H_T(Ay_{T-1} + Cx_T + b_T) - 2(Ay_{T-1} + Cx_T + b_T)'h_T + E_{T-1}u_T'H_Tu_T + c_T$$

by differentiating with respect to the vector x_T,

$$(7)\quad \frac{\partial W_T}{\partial x_T} = 2C'H_T(Ay_{T-1} + Cx_T + b_T) - 2C'h_T = 0$$

The solution of (7) yields the optimal policy for the last period,

$$(8)\quad \hat{x}_T = G_Ty_{T-1} + g_T$$

where

$$(9)\quad G_T = -(C'H_TC)^{-1}(C'H_TA)$$

$$(10)\quad g_T = -(C'H_TC)^{-1}C'(H_Tb_T - h_T)$$

Equation (8) is the optimal feedback control equation, showing that the optimal policy for period T is a linear function of all the variables y_{T-1} which will affect the outcome y_T through system (3). It also shows the multiperiod nature of our solution. The value of $\hat{x}_T$ cannot be set in period 1 since it depends on y_{T-1} which is not yet observed, but the optimal $\hat{x}_T$ can be specified as a function of y_{T-1}, and this function will be needed for us to determine the optimal policies for earlier periods.

To obtain the optimal expected cost for the last period, conditional on the data y_{T-1}, we substitute (8) for x_T in (6),

$$(11)\quad \hat{W}_T = y_{T-1}'(A+CG_T)'H_T(A+CG_T)y_{T-1} + 2y_{T-1}'(A+CG_T)'(H_Tb_T - h_T) + (Cg_T+b_T)'H_T(Cg_T+b_T) - 2(Cg_T+b_T)'h_T + c_T + Eu_T'H_Tu_T$$

Consider now the problem for one more period T−1. This two-period problem involves the choice of two policies, x_T and x_{T-1}, but we have already found an optimal policy $\hat{x}_T$ for the last period as a function y_{T-1}. The only remaining problem is to choose x_{T-1} optimally. The choice of x_{T-1} will affect y_{T-1}, but, whatever y_{T-1} will be, our previous solution guarantees that x_T shall be chosen optimally. This logic carries over to many periods. At the beginning of any period t, we are concerned only with the choice of x_t provided that we have obtained the optimal solution for the remaining periods from t+1 on. That solution will depend on y_t which the current decision x_t will influence, but whatever the outcome y_t may be, we know that all future decisions $x_{t+1}, \ldots, x_T$ will be optimally chosen. By the principle of optimality of Bellman, we minimize only with respect to x_{T-1}

$$(12)\quad E_{T-2}[y_{T-1}'K_{T-1}y_{T-1} - 2y_{T-1}'k_{T-1} + a_{T-1}'K_{T-1}a_{T-1} + \hat{W}_T]$$

where the welfare cost includes the contribution from the current period T−1 plus the optimal cost from period T on. Substituting (11) for $\hat{W}_T$ into (12) will yield

$$(13)\quad W_{T-1} = E_{T-2}[y_{T-1}'H_{T-1}y_{T-1} - 2y_{T-1}'h_{T-1} + c_{T-1}]$$

where

$$(14)\quad H_{T-1} = K_{T-1} + (A+CG_T)'H_T(A+CG_T)$$

$$(15)\quad h_{T-1} = k_{T-1} + (A+CG_T)'(h_T - H_Tb_T)$$

$$(16)\quad c_{T-1} = a_{T-1}'K_{T-1}a_{T-1} + (Cg_T+b_T)'H_T(Cg_T+b_T) - 2(Cg_T+b_T)'h_T + c_T + Eu_T'H_Tu_T$$

The solution for the T-period problem is complete if one observes that the expression (13) to be minimized with respect to x_{T-1} has the same form as expression (5). One can thus repeat the process from

equations (5) to (16), with the subscript T replaced by T−1, and so forth. In summary, optimal control consists of choosing the instrument x_t as a linear function $G_t y_{t-1}+g_t$ of the variables y_{t-1} of the previous period, as in equation (8). The matrices of coefficients G_t are determined, together with H_t, by solving equations (9) and (14) alternatively, backward in time from t=T, and with initial condition $H_T=K_T$. The matrices G_t and H_t having been obtained, the vectors g_t, together with h_t, are determined by solving equations (10) and (15) alternatively, backward in time from t=T, and with initial condition $h_T=k_T\equiv K_T a_T$.

Of course, in the beginning of period 1, one needs to act only on x_1, but the above derivation shows that the optimal $\hat{x}_1$ depends on G_1 and H_1, which in turn depend on the future H_t and G_t, t=2, . . . , T. By the multiperiod nature of our problem, future policies have to be taken into account in the determination of the optimal policy for the first period. Our calculations will also yield the expected welfare cost associated with the optimal policy from periods 1 to T, given the initial condition y_0. It is given by equation (11), with T replaced by 1, thus being a quadratic function of y_0. In the calculation of $\hat{W}_1$, we need c_1, which can be obtained by solving equation (16) backward in time.

If the matrices K_t in the welfare function are all equal to K, the solution for G_t and H_t may reach a steady state, for t smaller than a certain value, that will satisfy

$$G = -(C'HC)^{-1}C'HA; \tag{17}$$

$$H = K + (A + CG)'H(A + CG) \tag{18}$$

Since (18) can be written as an infinite series

$$H = K + (A + CG)'K(A + CG) + (A + CG)'^2K(A + CG)^2 + \dots \tag{19}$$

the steady state will exist if, and only if, the series converges. Convergence occurs if all the characteristic roots of the matrix $(A+CG)$ are smaller than one in absolute value. Even when G_t and H_t do reach a steady state, g_t and h_t will usually not do so if $k_{t-1}(\equiv Ka_{t-1})$ and b_t are changing through time, as can be seen from equation (15). If G_t is constant for t=1, 2, 3, . . . , the optimal policies $\hat{x}_t$ should respond to the data y_{t-1} in the same way, but the intercepts g_t will change because the targets a_t and the effects b_t of the uncontrolled exogenous variables are changing—optimal policies should take these changes into account.

II. Problems Associated with Instrument Instability for Deterministic Systems

The questions raised by Holbrook are for a deterministic model, i.e., model (1) with $u_t=0$. It may be of interest to consider this special case first. Note that the above formulation of, and solution to, the optimal control problem applies easily to this case—simply erase the expectation signs and set u_t equal to zero. The linear feedback control equations $x_t=G_t y_{t-1}+g_t$ and the computations of G_t and g_t remain the same as in the stochastic case.

Before analyzing any questions, one may wish to define instrument instability. Holbrook did not give a precise definition. By this term, one might mean (a) that the optimal time paths of x_t are explosive and/or (b) that they contain oscillations. If x_t satisfies a system of linear difference equations, (a) will occur when some roots of the system are larger than unity in absolute value; (b) can occur if some roots are complex and/or negative.[1] I suppose that the degrees of (a) and (b), namely, how explosive and how large the oscillations, should also matter. Thus the

[1] Proofs of these well-known propositions can be found in Chow (1968).

absolute values of the roots, be they real or complex, will matter. If one specifies that the first differences of economic variables satisfy a linear model like (1), he might reasonably require, in his definition of stability, that the instruments x_t (also in first differences) be not explosive, or not very explosive. But if the model explains the levels of economic variables which are expected to grow, some explosiveness in the instruments should not be considered unstable. In short, instrument instability requires one to examine the explosiveness and/or the extent of oscillations of the optimal time paths of the instruments x_t.

Will instrument instability exist for a particular system? To answer this question, one simply studies the optimal path of x_t, which, in the notation of (3), is imbedded in the vector y_t. When optimally controlled, y_t will satisfy (in the deterministic case)

$$(20)\quad \begin{aligned} y_t &= Ay_{t-1} + Cx_t + b_t \\ &= (A + CG_t)y_{t-1} + b_t + Cg_t \end{aligned}$$

and the behavior of such a system can easily be analyzed, especially when G_t reaches the steady state G. For $G_t=G$, and denoting b_t+Gg_t by $\bar{b}_t$, say, the solution to (20) is

$$(21)\quad \begin{aligned} y_t = (A + CG)^t y_0 &+ \bar{b}_t \\ &+ (A + CG)\bar{b}_{t-1} + \dots \\ &+ (A + CG)^{t-1}\bar{b}_1 \end{aligned}$$

in which the homogeneous part $(A+CG)^t y_0$ is explosive or oscillatory, depending on whether the roots are large in absolute value or complex.

One might ask, if the system omitting the influence of x_t, i.e.,

$$(22)\qquad y_t = Ay_{t-1} + b_t$$

is explosive (damped), will the system including the instruments x_t which are set optimally, i.e., system (20), be explosive (damped) also? Either case can happen, but to dampen an otherwise explosive system by optimal control is more likely than to change an otherwise stable system to an explosive one. After all, a purpose of control may be to dampen an explosive system. From equation (9) for G_t (or from equation (17) for G), each column of the optimal feedback control matrix G_t can be interpreted as coefficients of a regression of the corresponding column of the matrix A on columns of the matrix $-C$. In other words, if the first column of A, say A_1, are observations on a dependent variable, and the columns of $-C$ are observations on the explanatory variables, then the first column of G_t, say $G_{t,1}=-(C'H_tC)^{-1}C'H_tA_1$, are coefficients in the regression $A_1 = -CG_{t,1}+R_{t,1}$ obtained by the method of Aitken's generalized least squares. The purpose of regression is to make the columns of residuals

$$(R_{t,1}R_{t,2}\dots) = R_t = A + CG_t$$

small—in fact, for column 1, it is to minimize the weighted sum of squares of the residuals $R'_{t,1}H_tR_{t,1}$. Therefore, if some roots of A are larger than one in absolute value, the roots of the matrix $R_t=A+CG_t$ of regression residuals may not be, in which case the system under control is damped.

How would it be possible for the matrix A to be stable, but the matrix $A+CG$ (omitting subscript t) to be unstable? It is possible, when the weighting matrix H, which is derived from the matrix K of welfare weights by equation (14), is uneven in assigning weights to different residuals. As an example, consider the simple system given in the third paragraph of the introduction of this paper. In the notation of (2) and (3), this system is

$$\begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} 0 & .6 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} .4 \\ 1 \end{bmatrix} x_t + \begin{bmatrix} 10 \\ 0 \end{bmatrix}$$

The matrix A has multiple roots of zero. If the purpose of control is to steer y_t to target, but to ignore the behavior of x_t, the 2x2 matrix K is diagonal with unity as its leading element, and

$$(23) \quad G_T = -\left[(.4 \quad 1)\begin{pmatrix}1 & 0\\0 & 0\end{pmatrix}\begin{pmatrix}.4\\1\end{pmatrix}\right]^{-1} \cdot (.4 \quad 1)\begin{pmatrix}1 & 0\\0 & 0\end{pmatrix}\begin{pmatrix}0 & .6\\0 & 0\end{pmatrix} = -(.16)^{-1}(0 \quad .24) = (0 \quad -1.5)$$

$$(24) \quad A + CG_t = \begin{pmatrix}0 & .6\\0 & 0\end{pmatrix} + \begin{pmatrix}0 & -.6\\0 & -1.5\end{pmatrix} = \begin{pmatrix}0 & 0\\0 & -1.5\end{pmatrix}$$

$$(25) \quad H_{T-1} = K + (A + CG_T)'K(A + CG_T) = K = \begin{pmatrix}1 & 0\\0 & 0\end{pmatrix}$$

Therefore, the values for G_T and H_{T-1} given in (23) and (25) are steady-state values, and we have an explosive root -1.5 in the residual matrix $A+CG_T$.

One can now state several relevant propositions concerning optimal control that are based on the theory of Section I. First, it is possible to steer certain endogenous variable(s) exactly to targets, under appropriate conditions to be specified below, by solving the optimization problem of Section I that employs the quadratic welfare function (4), and assigning positive costs only to the deviations of the variable(s) selected. Second, if the number of target variables (the number of nonzero diagonal elements in K) is equal to the number of instruments, the minimization problem of equation (5) will be solved with $(y_T - a_T)'K(y_T - a_T) = 0$, and so on for equation (13), etc., and the variables will be on target exactly—assuming that the rank of C is equal to the above number. Third, if the number of target variables is larger than the number of instruments, the former variables will not reach targets exactly, and their deviations from targets will depend on the welfare weights in K assigned to them.[2]

Fourth, instrument instability, insofar as it pertains to some characteristic of the time path x_t, can exist no matter whether the welfare weights are assigned only to the originally endogenous variables, or also to the instruments (imbedded in y_t). Holbrook treated only the former case, in fact, a rather special case of the former when the numbers of target variables and of instruments are both equal to one. Fifth, once the last point is recognized, as it is from the theory of Section I, one can trade off the stability characteristics not only between the truly endogenous variables and the instruments, but also among the instruments themselves (and among the endogenous variables themselves, of course). All these can be accomplished by changing the diagonal elements of the matrix K. For example, if total government expenditures and money supply are two instruments, giving more weight to the former (and specifying its target path to be smooth) will mean more stability for it, as compared with the latter instrument, when the weights for the endogenous variables remain unchanged.

Once the basic ideas of optimal control are understood, one might not wish to take the concept of instrument instability very seriously. What matters are the dynamic characteristics of the time path in question, be it a dependent variable or an instrument. Moreover, within the framework of Section I, a good measure of the dynamic performance of the variable is the sum of its squared deviations from the target path over time. If the variable fluctuates violently, and the target path is

[2] If the number of target variables is smaller than the number of instruments, there will be more than one way to achieve the targets exactly, and the computation of the optimal control matrices G_t can be performed by using a generalized inverse for $(C'H_tC)$, where the rank of H_t and thus of the entire matrix, is smaller than the number of instruments.

smooth, its performance is poor by this measure. If the variable increases rapidly, and the target calls for no increase or a small increase per period, its performance is poor by this measure. Having solved the problem of optimal control, one can easily compute the optimal path of y_t (including x_t as a subvector) using equation (20) and thus the sums of squared deviations of all variables from targets, and, in fact, a weighted sum of these sums as a measure of welfare cost.

Having discussed instrument instability, we have also provided answers to the problems of setting objectives, ascertaining the existence of good solutions (by simply calculating $\hat{W}_1$), finding optimal solutions, and, to a lesser extent, welfare tradeoffs between target variables, possibly including instruments. The last problem and the remaining problems mentioned in the introduction will be treated in Section IV. In the next section, modifications required by the stochastic nature of the problem will be discussed.

III. Modifications for Stochastic Systems

For a stochastic system, one retains the random disturbance u_t in equation (3) and the expectation sign in equation (5), etc. The objective of control, in the framework of Section I, is to minimize the expectation of a weighted sum of squared deviations of selected variables from targets. In this section we will point out several implications of including the random disturbance in the dynamic system (3).

First to be noted is the fact that the optimal control equations $\hat{x}_t=G_t y_{t-1}+g_t$ remain exactly the same as in the deterministic case. However, there is an important difference in that the time path of the optimal $\hat{x}_t$ is now stochastic, since y_{t-1} is stochastic, whereas both $\hat{x}_t$ and y_{t-1} were deterministic before. In other words, when $u_t=0$, it was possible to calculate, in period 1, the optimal $\hat{x}_t$ for all t, since all future y_t were to be determined by equation (21). With stochastic disturbances present, future y_t becomes stochastic, and optimal $\hat{x}_t$ can be determined numerically only after y_{t-1} are observed.

It is fruitful to decompose the stochastic time series y_t generated by the system (3), when x_t is optimally controlled, into two parts. The first part $\bar{y}_t$ is deterministic and the second y_t^* is stochastic. When x_t follows the optimal policy $\hat{x}_t=G_t y_{t-1}+g_t$, system (3) will become, with G_t reaching the steady state G, and with $\bar{b}_t$ denoting b_t+Cg_t:

$$(26)\qquad \begin{aligned} y_t &= (A+CG)y_{t-1}+\bar{b}_t+u_t \\ &= Ry_{t-1}+\bar{b}_t+u_t \end{aligned}$$

After repeated substitutions of y_{t-1} in (26) by $Ry_{t-2}+\bar{b}_{t-1}+u_{t-1}$, and y_{t-2} by $Ry_{t-3}+\ldots$, etc., one easily finds that

$$(27)\qquad \begin{aligned} y_t = R^t y_0 &+ (\bar{b}_t+R\bar{b}_{t-1}+\ldots+R^{t-1}\bar{b}_1) \\ &+ u_t+Ru_{t-1}+\ldots+R^{t-1}u_1 \end{aligned}$$

The first line on the right-hand side of (27) is the mean of the process, to be denoted by $\bar{y}_t$. It is identical with the time path (21) of the deterministic system of Section II, with $\bar{y}_0=y_0$. The second line of (27) is the deviation from mean, or the random part of the process, to be denoted by y_t^*. It is generated by the process

$$(28)\qquad y_t^* = Ry_{t-1}^* + u_t$$

If one were interested in studying instrument instability for this dynamic stochastic system under control, he would have to define "instability" in terms of certain dynamic characteristics of a stochastic time-series, and to ascertain these characteristics for the time series y_t given by (27). Once y_t is decomposed into $\bar{y}_t$ and y_t^*, there are ways to characterize its explosiveness and cyclical fluctuations, and to derive the measures of these characteris-

tics for system (27).[3] We will not pursue the subject in this paper because instrument instability is not considered to be a subject of great interest in itself and because an adequate treatment would be too lengthy here.[4]

As it was pointed out in Section II, instrument instability should be taken less seriously than, or should be replaced by, the weighted sum of squares of deviations of y_t from the targets a_t, where y_t incorporate both instruments and truly endogenous variables. In the stochastic case, it is the expected weighted sum of squares given by (4) for any policy, or by $\hat{W}_1$ for the optimal policy. Equation (11) for $T=1$ and equation (16) can be used to compute $\hat{W}_1$ for the stochastic case. Note that, in the deterministic case, all $Eu_t'H_tu_t$ in the calculation of $\hat{W}_1$ would be zero. Thus the (optimal) expected welfare loss in the stochastic case equals the loss for the deterministic case, i.e., for the time-series $\bar{y}_t$, plus the expected loss due to the random disturbances, i.e., from the time-series y_t^* of equation (28). The former component is due to deviations of $\bar{y}_t$ from a_t, while the latter component is due to the deviations of y_t from $\bar{y}_t$, namely, the deviations y_t^*.[5] This decomposition deserves several comments. First, even if the number of instruments equals the number of target variables, one cannot expect the stochastic time-series y_t to be on target exactly because of random disturbances, and the expected squared deviations will contribute to the expected welfare loss, as measured by the second component of $\hat{W}_1$—here the first component is zero because the deterministic path $\bar{y}_t$ is exactly on target. Second, when the number of target variables exceeds the number of instruments, as it is more often the case, both components of $\hat{W}_1$ will be positive. Third, calculations based on the simple, aggregative macro-economic model of Chow (1967) have shown that, for a number of reasonable welfare functions, the component of $\hat{W}_1$ due to the stochastic disturbances is invariably much larger than the deterministic component, indicating the importance of considering the stochastic aspect of the problem in the measurement of welfare cost.[6] Thus, although the introduction of random disturbances does not change the optimal control equations, it materially affects the nature of the time series generated by the controlled system, and increases the optimal expected welfare cost.

Perhaps it is felt that recognizing the random disturbances is to recognize only a part of the problems associated with a stochastic system. Another important aspect is the uncertainty of the exact values of the parameters A and C of the system (3). When A and C are estimated statistically, the errors in the estimates have to be taken into account in the design of optimal policies, as they would affect both the optimal control equations and the associated expected welfare loss. This subject is treated in Chow (1972c) and will not be pursued further here. For an approach to control problems with non-linear systems, the reader is referred to Michael Athans.

IV. Other Problems of Economic Policy

First, let us consider the problem of ascertaining the tradeoff possibilities between different target variables. The optimal control solution will minimize the welfare loss by appropriately weighting the respective sums of squares of deviations of these variables from targets ac-

[3] These have been developed and discussed at length in Chow (1968), (1970), Chow and Richard Levitan (1969a, b).

[4] The treatment would be straightforward by methods set forth in the references cited in the preceding footnote.

[5] For alternative developments of this decomposition, see Chow (1972a, b).

[6] These calculations can be found in Chow (1972b).

cording to the elements of the matrix K. In the deterministic case, the solution provides a given value for each target variable for each time period. In the stochastic case, the mean values of the target variables are given by $\bar{y}_t$, as well as their variances and covariances by the variances and covariances of y_t^*. Thus, the sum of squared deviations over time of each variable from its target path, or the expectation thereof in the stochastic case, is provided by the optimal solution. For example, the average squared deviation over time for the rate of unemployment (from zero) may turn out to be the square of 6 percentage points, and the average squared deviation for the rate of inflation (from zero) may be the square of 4 percent per year. To obtain another point in the unemployment-inflation tradeoff, say with a reduction of unemployment, one simply changes the diagonal elements of the matrix K to make the deviation of unemployment more costly relative to that of inflation, and calculates another optimal solution. The latter may turn out to be a combination of 5 percent unemployment and 5 percent inflation.

There appear to be advantages of this approach for ascertaining the tradeoff possibilities between different objectives over the approach by simulations as it is currently practiced. First, we are forced, in the optimal control framework, to make some attempt to rank the different sets of time paths for the target variables to be generated by different policies. Second, the ranking is done not only for one sample from each policy but for the average over samples as derived analytically by taking mathematical expectation. Often, simulation studies ignore the stochastic disturbances in the system. If they do not, averaging over many samples generated by random drawings from the distribution of the disturbances will be required, and the errors in estimating the true mathematical expectation of the squared deviations are often difficult to assess. Furthermore, and this is the third point, without doing some optimization as a guide, one would have to try many different policies in a simulation study. This not only aggravates the difficulties of the second point, but more importantly, one may never discover the very good policies. From the viewpoint of tradeoffs, one may miss the policies that would lower both the rates of unemployment and inflation.

A purist may object to the method of ascertaining the tradeoffs as described above. If one were to take his quadratic welfare function seriously, and obtain an optimal solution with 6 percent unemployment and 4 percent inflation for some year, this must be the best that he can do. There is no reason for exploring whether a combination of 5 percent and 5 percent is feasible, because, even if it were, the policy maker would not prefer it as it is not optimal according to his solution. There are at least two answers to this. First, another policy maker may have a different welfare function, and may prefer a 5-5 combination to a 6-4, and it would be useful to exhibit various possibilities. Second, a quadratic welfare function has many well-known limitations. It may be a reasonable approximation to the preference of the policy maker only in a certain region of the target-variable space, say near the target point a_t of (3, 2) for unemployment and inflation and to the right and above that point. If the optimal solution is far from that region, the policy maker will want to write another quadratic approximation to his preferences to see whether the result will turn out to be better. This will amount to the procedure which we have recommended, with the amendment that the target point itself is subject to change, besides the matrix K.

Another problem of interest is the evaluation of policy proposals. For ex-

ample, would increasing money supply by 5 percent per year be a good policy? Such a proposal is not a complete specification of a policy. What are to be assumed for the other instruments? Are they to follow an optimal feedback rule while money supply is to be increased at a constant rate, or are they to maintain constant rates of change also? In the latter case, at what rates? By what criteria is the dynamic performance of the economy to be judged? Using the framework of optimal control, I have specified target paths for certain econlmic variables generated by the model of Chow (1967) and obtained the suboptimal policies under the restriction that the two instruments, money supply and government expenditures, are to grow at constant rates—the rates being determined suboptimally for each welfare function specified. I have then compared the welfare losses associated with these policies with those for the optimal policies, and have found that the former are from 40 to 80 percent larger than the latter for the given econometric model. The results are reported in Chow (1972b), and they constitute an evaluation of policy proposals by the method of optimal control.

For both the evaluation of policy proposals and the appraisal of historical policies, it is necessary to use a dynamic model in order to take into account the effects of a decision in one period on the outcomes of later periods. It is also necessary to treat the decision of one period in conjunction with the decisions of other periods. It would be inadequate to evaluate an action solely by the consequence in the same period, or to ignore the structure of intertemporal decisions.[7] The desirability of using a welfare function as a yardstick and of using an optimal control solution as an alternative for comparison require no further comment here. Furthermore, for the appraisal of historical record, it would be desirable to describe the behavior of policy makers by some decision rule, rather than by the numerical outcomes of their actions. This suggestion is entirely consistent with the recent research interest in explaining government policy variables by systems of equations. These equations are useful for forecasting purposes, and the explanation of government behavior quantitatively is in itself important. For the purpose of historical appraisal, this suggestion is consistent with the fundamental proposition in the study of decision under uncertainty that a decision should be judged by the expected utility of the outcome when it is made, and not by the utility of the outcome that actually occurs since the latter is affected by chance elements. In a deterministic world, one could use a nonstochastic economic model, and compare the time paths of the economic variables generated by the historical values of the instruments, treated as constants, and the time paths generated by alternative values of the instruments, also treated as constants. When there are random disturbances, the historical values might be sample outcomes of certain rules and should not be judged as if the policy makers had specified these quantities as constants independent of the random events that had occurred. By incorporating empirically estimated dynamic relationships describing government policies into an econometric system, as we did for the optimal policies, one can study the dynamic performance of the

[7] Without the benefits of the tools of optimal control, early studies in the evaluation of macro-economic policies ran into these difficulties. For example, Martin Bronfenbrenner, in comparing rival monetary rules, assumed that the effect of a change in money supply would work out in the period of one year. The study of Franco Modigliani, in comparing for each period the historical money supply with some "target" money supply which the author specified, also failed to consider the dynamic consequences of a decision. Furthermore, the target money supply could hardly have been an optimal path since intertemporal decision was not considered.

economy as it was historically regulated. Comparison of the expected welfare loss and other properties, with those of an optimal policy, or of an alternative policy proposal, can then be made. A crude beginning based on this approach was described in Chow (1972b).

As the last problem, consider the measurement of the effectiveness of monetary versus fiscal policies. It is generally recognized that comparing the coefficients of money supply and government expenditures in a regression of national income on these variables will leave many questions unanswered. Is national income the only variable of interest? How can coefficients be compared without giving considerations to the costliness of changing the policy variables? How about the dynamic nature of the problem? The lagged effects of the instruments will have to be judged, but not by simply inspecting the sets of coefficients in a distributed-lag relationship between national income and the instruments, because a relevant comparison requires letting each set of policy instruments perform at its best under a dynamic and stochastic setting, namely, letting them follow some optimal control rules. This last point applies equally well to the study of effectiveness of different policy instruments by the method of simulations using dynamic econometric models. Not only is the estimation of expected welfare associated with each policy by way of stochastic simulations involving many variables for many periods costly and subject to errors, as pointed out before, but it is virtually impossible to design rules for these instruments which will show their true "effectiveness" without some optimization procedure as described in this paper.

I hope to have demonstrated that the concepts and tools of this paper are extremely useful and important for the study of economic policy, and that this subject can be greatly advanced by an approach that is fundamentally multivariate, dynamic, forward-looking, and stochastic in conception. The theory of macro-economic policy need no longer be built on comparative statics under the assumption of certainty, but rather should be concerned with problems of comparative dynamics in a stochastic setting.[8] In the empirical study of economic policy, the tools of optimal control should not be regarded as capable of yielding one final solution to policy, but rather should be applied with flexibility as a means of uncovering the preferences of decision makers, the reasonably good possibilities open for policy, and thus the dynamic characteristics of the economy.

REFERENCES

M. Athans, "The Discrete Time Linear-Quadratic-Gaussian Stochastic Control Problem," *Annals Econ. Soc. Measurement*, Oct. 1972, *1*, 449–91.

R. E. Bellman, *Dynamic Programming*, Princeton 1957.

M. Bronfenbrenner, "Statistical Tests of Rival Monetary Rules," *J. Polit. Econ.*, Feb. 1961, *69*, 1–14.

E. C. Brown, "The Static Theory of Automatic Fiscal Stabilization," *J. Polit. Econ.*, Oct. 1955, *63*, 427–40.

G. C. Chow, "Multiplier, Accelerator, and Liquidity Preference in the Determination of National Income in the United States," *Rev. Econ. Statist.*, Feb. 1967, *49*, 1–15.

———, "The Acceleration Principle and the Nature of Business Cycles," *Quart. J. Econ.*, Aug. 1968, *82*, 403–18.

———, "Optimal Stochastic Control of Linear Economic Systems," *J. Money, Credit, Banking*. Aug. 1970, *2*, 291–302.

———, (1972a) "Optimal Control of Linear Econometric Systems with Finite Time Horizon," *Int. Econ. Rev.*, Feb. 1972, *13*, 16–25.

[8] Some early contributions to the theory, including E. Cary Brown, p. 430, for example, recognized the importance of comparative dynamics, but the tools were not available.

———, (1972b) "How Much Could be Gained by Optimal Stochastic Control Policies," *Annals Econ. Soc. Measurement*, Oct. 1972, *1*, 391–406.

———, (1972c) "Effect of Uncertainty on Optimal Control Prices," *Int. Econ. Rev.*, forthcoming.

——— and **R. E. Levitan**, (1969a) "Spectral Properties of Non-Stationary Systems of Linear Stochastic Difference Equations," *J. Amer. Statist. Assn.*, June 1969, *64*, 581–90.

——— and ———, (1969b) "Nature of Business Cycles Implicit in a Linear Economic Model," *Quart. J. Econ.*, Aug. 1969, *83*, 504–17.

R. S. Holbrook, "Optimal Economic Policy and the Problem of Instrument Instability," *Amer. Econ. Rev.*, Mar. 1972, *62*, 57–65.

F. Modigliani, "Some Empirical Tests of Monetary Management and of Rules versus Discretion," *J. Polit. Econ.*, June 1964, *72*, 211–45.

The Interaction Between Federal Credit Programs and the Impact on the Allocation of Credit

By Rudolph G. Penner and William L. Silber*

Federal credit programs and policies aimed at increasing the amount of resources devoted to specific types of "socially desirable" expenditures, such as housing, have expanded considerably during the past five years. In general, these policies and specific programs have costs. They also interact with one another in an, as yet, undefined way. Our objectives in this paper are to describe the determinants of the relative effectiveness of the different government credit programs and policies in implementing the flow of funds (and resources) into a particular sector. We also analyze the interaction between different credit programs. Housing will be used as the specific example because it is currently high on the priority list of government policy and, more importantly, there are a myriad of policies and programs aimed at promoting housing.

Three qualifications are in order at this point. We include in the term "costs" the total *direct* cost to the government and private sector arising from a particular policy. We do not compare the indirect costs of a program such as the possible reduction in the efficiency of the financial sector or any excess burdens of a policy in the welfare sense. We also do not question the government's concern with housing. The government intervenes and subsidizes a sector of the economy when it believes that real social benefits exceed private benefits or real social costs are less then the private sector's evaluation of real costs. We assume this to be the case in housing. Finally, we do not analyze the costs and benefits of diverting the flow of resources via credit programs as opposed to attaching subsidies to the real side of, say, housing activity. Most government housing programs are aimed at diverting the flow of credit. Our concern is with the interaction between and impacts of these programs.

In Section I of the paper we provide the conceptual and analytic framework. In Section II the actual programs are described in detail and are then placed within the context of the theoretical model. The paper concludes with the policy implications of the analysis.

I. Theoretical Framework

We start with a set of simplifying assumptions which eliminate certain complications from the analysis without severely limiting the generality of the conclusions. First, we measure the impact of housing credit programs by the effect on the flow of mortgage credit and the impact on the terms of credit. While such effects on credit are likely to have an impact on housing activity (since mortgage

* Professor of economics, University of Rochester, and professor of economics, Graduate School of Business, New York University, respectively. Silber wishes to acknowledge the financial support of the Federal Home Loan Bank System. In addition, secretarial assistance was generously provided by the Federal Home Loan Bank of New York. The views expressed are not necessarily shared by the Federal Home Loan Bank System. We wish to thank the following for their comments: Ernest Bloch, Alan Meltzer, Bruce Ricks, and Lawrence Ritter.

flows and rates are related empirically to housing activity),[1] this relationship is not essential for our purposes. Second, the analysis proceeds as though the effectiveness of programs is related solely to their ability to lower the mortgage rate of interest paid by the borrower or increase the flow of mortgage credit. This ignores the fact that most credit programs apply to particular types of mortgages or to particular groups in the population (for example, veterans, poor people, etc.) and that different groups may have different elasticities of demand for housing with respect to the mortgage rate. A detailed comparison of elasticities is beyond the scope of this paper. Third, for simplicity, the mortgage rate of interest is specified as the only determinant of the supply and demand for mortgage funds (implicitly including the amortization period and the loan-value ratio and their impact on mortgage supply and demand). Finally, we also assume that various government programs are well managed and pursue an optimum strategy in trying to achieve their stated goals. For example, we ask what conditions determine whether a well-run secondary market program has a positive impact on the supply of mortgage credit. We do not ask whether the actual secondary market program is well managed.

Given these assumptions, it is useful to classify credit programs as follows:

1) Policies in this category are designed to affect the interest rate paid by mortgage borrowers without changing the risk characteristics of such securities. Two very different techniques are used in the pursuit of this goal: (a) Programs in this subgroup provide *explicit* interest subsidies which introduce a wedge between the mortgage rate received by the lender and that paid by the borrower, usually lowering the latter and raising the former. (b) Policies in this subcategory lower *both* the borrower and lender rates by inducing lending institutions (for example, via portfolio regulations on savings and loan associations) to raise their supply of mortgage funds at all rates of interest. The cost of programs under 1a) are explicit (the subsidy) while cost of programs under 1b) are implicit. The incidence of the costs under 1b) will be discussed below.

2) In this category the programs are designed to first change the risk characteristics of mortgages so that they become more desirable assets to lenders (for example, a reduction in lenders' risk by government insurance and guarantee programs or by the creation of a secondary market). This reduction in risk then raises the supply of mortagage funds and, presumably, lowers both borrower and lender rates. In these programs costs can be both explicit and implicit as we will see below.

This classification of credit programs permits a formal analysis of their relative effectiveness. Policies under 1a) and 1b) will be called wedge-type programs and portfolio-restriction type policies, respectively, and will be discussed first. Programs listed under 2) will be called mortgage characteristic programs, and will be discussed last.

A. *Determinants of Policy Effectiveness*[2]

For any given demand curve for mortgage funds, a wedge-type subsidy on all interest paid attracts more additional funds to the market the more interest elastic is the supply curve of mortgage funds. A highly elastic supply curve exists when mortgages are good substitutes for

[1] For empirical evidence on the impact on mortgage credit and mortgage rates on housing, see Eugene Brady; Ray Fair, ch. 8; Craig Swan.

[2] Work in this area has been done by George Break, Jack Guttentag and D. C. Rao and Ira Kaminow.

other assets in investor portfolios (i.e., a small increase in the relative rate of return on mortgages induces portfolio substitution in favor of mortgages and a corresponding increase in supply).

This positive relationship between the degree of substitutability between mortgages and other securities and the effectiveness of wedge-type subsidies is illustrated in Figures 1(a) and 1(b). The mortgage rate, i_M, is on the vertical axis and the dollar volume of mortgages, $\$M$, is on the horizontal axis. The demand curve for mortgage funds, D, is negatively related to i_M. The supply curve, S, is generally a positive function of i_M. The elasticity of supply is determined by the degree of substitutability between mortgages and other securities. The supply curve shifts with a change in the risk characteristics of mortgages and also shifts when rates on "other securities" change as long as mortgages and other securities are substitutes or complements.

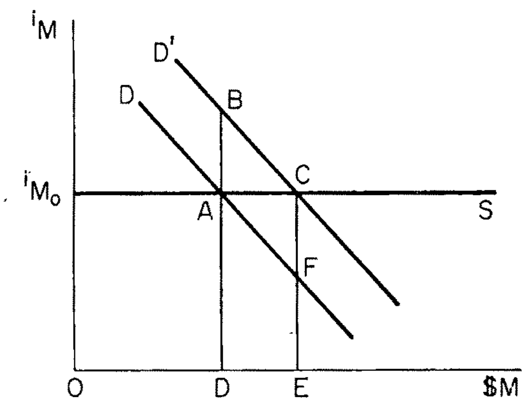


FIGURE 1a

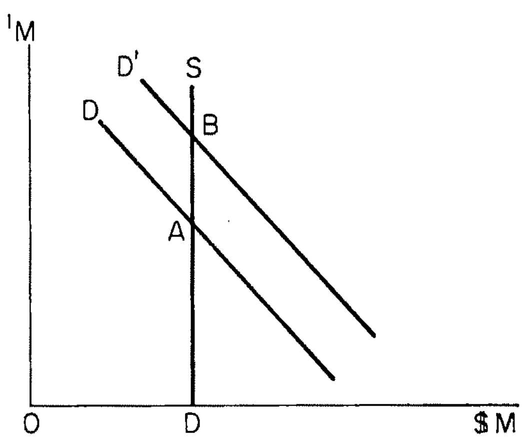


FIGURE 1b

In Figure 1(a), the supply curve of mortgage funds is infinitely elastic at rate i_{M_0}. This situation occurs when mortgages and other securities are perfect substitutes at the existing rate structure.[3] Introducing an interest subsidy to borrowers equal to AB raises the demand curve for mortgage funds to D'. This elicits an increased supply of mortgage funds in the amount DE. The new effective rate to the borrower is EF and the total supply of mortgage funds now stands at OE. Since mortgages and other securities are perfect substitutes, neither i_{M_0} (the rate received by lenders) nor the rate on other securities change, hence, there is no feedback to the supply curve of mortgage funds due to changes in other rates.

In Figure 1(b) the case of an infinitely inelastic supply curve is illustrated, corresponding to the zero substitutability case. Here the effective rate to the borrower remains at DA, the rate received by lenders increases by AB and mortgage funds remain at OD. There is no feedback to the supply curve of mortgages from other securities markets due to the independence condition.

The case of some, but not infinite, substitutability between mortgages and other securities produces results that lie between the extremes in Figures 1(a) and 1(b). In other words, with a positively sloped mortgage supply curve the increase in mortgage funds is less than DE but more

[3] See James Tobin for a discussion of the determinants of substitutability between securities and the definition of perfect substitutes.

than zero and the decline in the effective rate to borrowers is less than CF but more than zero (this result can include a leftward shift in the mortgage supply curve due to a rise in rates on other securities within a general equilibrium framework).

It is also important to note that in this case the source of the subsidy funds has an impact on the general equilibrium solution. If the subsidy funds are raised by floating bonds in the other securities market, this shifts the demand curve for funds in that market upward, raising the rate on other securities, and hence shifting the supply curve for mortgages to the left, thereby partially offsetting the impact of the subsidy. On the other hand, if the subsidy funds come from general tax revenues (or in the extreme from taxes which reduce consumption only), the total supply of funds to the capital market is augmented and there need not be an offset to the subsidy.

The above discussion applies to a uniform subsidy paid on all mortgage interest. There is presently no program of this type; rather, different subsidies are paid to different groups in the population. Some are explicitly aimed at lower income groups and others at middle income groups. The details of such programs will be discussed in Section II.

We now turn to portfolio-restriction type programs. Here we can show that the *lower* the degree of substitutability between mortgages and other securities, the *larger* will be the inflow of funds to the mortgage market (and the larger will be the fall in the mortgage rate) as a result of the restrictions. Note that this substitutability condition is just the reverse of that which made wedge-type programs highly successful.

In Figures 2(a) and 2(b) we have two supply curves of mortgages, S_2, which is the supply curve of the institution (or sector) that is to be subject to restrictions

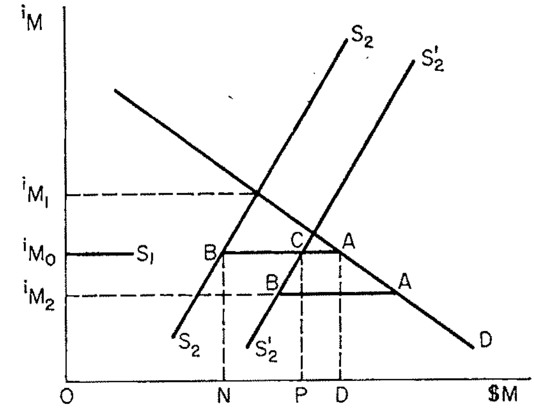


FIGURE 2a

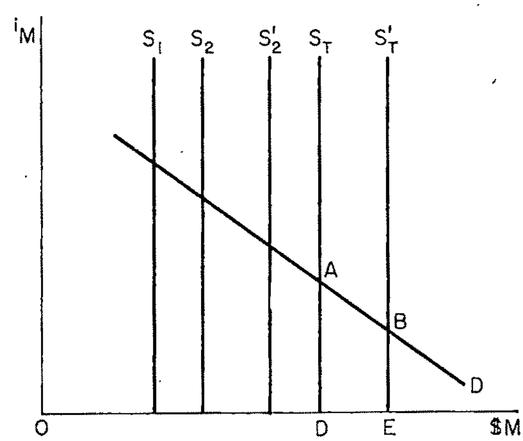


FIGURE 2b

and S_1, which is the supply curve of the unrestricted sector. The sum of the two is the total supply curve of mortgages. In Figure 2(a) the supply curve S_1 is infinitely elastic at rate i_{M_0}, in other words, for the unrestricted sector mortgages and other securities are perfect substitutes at that rate. Without this sector of the market the mortgage rate would be i_{M_1}, (the intersection of S_2 and D). At that rate, however, the unrestricted sector buys mortgages until the market rate is driven down to i_{M_0} at which point total mortgage funds supplied equals OD. The effective total supply curve is S_2BA, with institution 2

buying ON mortgages and the rest of the market buying ND.[4]

The imposition of portfolio restrictions on institution 2, such as via a law which restricts its security holdings to mortgages only, produces a rightward shift in the supply curve of institution 2, say, to S_2'. This rightward shifts occurs as long as the institution initially has a diversified portfolio, i.e., as long as it is a risk averter. If all other participants continue to buy BA (equal to ND) then the mortgage supply curve would be $S_2'BA$ and the mortgage rate would be i_{M_2}. At that rate, however, the unrestricted sector sells mortgages (buying the other securities, institution 2 must sell). Equilibrium is reestablished when the rate is back at i_{M_0} with institution 2 holding OP mortgages and the unrestricted sector holding PD. Total mortgages supplied remains OD. In other words, when restrictions are imposed in a situation where, for the market as a whole, securities are perfect substitutes, there is no impact on the effective mortgage rate or on the supply of mortgage funds.

In Figure 2(b) the extreme opposite case is illustrated. The supply curve of each sector is vertical, implying zero substitutability between mortgages and other securities. The intersection of $S_T(=S_1+S_2)$ and the demand curve, D, produces an equilibrium rate DA. The imposition of restrictions shifts S_2 to S_2', the total supply curve shifts (by the same amount) to S_T' and mortgage funds increase to OE and the rate falls to EB. There is no offset to the increased supply of mortgages by S_2 due to a decreased supply in the unrestricted sector since both supply curves are interest inelastic. There is also no offset due to changes in the rate on other securities because of the independence condition.

[4] Note that the division of mortgage holdings between the two sectors is merely illustrative. The conclusions of this analysis are unaltered by a different distribution.

With some, but not infinite, substitutability between mortgages and other securities the results lie between those depicted in Figures 2(a) and 2(b). In other words mortgage supply rises by less than DE but by more than zero and the rate falls by less than it does in 2(b).

Portfolio restriction policies also impose implicit costs which partially offset their effectiveness, leading, in turn, to further government programs. In particular, forcing an institution to specialize in mortgages (either by explicit regulation or by imposing high reserve ratios against other assets) imposes a higher level of risk per expected rate of return on the institution (compared with unconstrained optimization). In effect the restrictions impose an implicit tax on the institution. As with any other tax, part of the burden (cost) will be shifted away from the institution's owners. For example, if the restricted institution's liabilities are poor substitutes for the liabilities of nonrestricted institutions (and other securities), a reduction in the rate of return on their liabilities does not produce a significant loss of funds. In this case the burden of the restrictions falls mainly on the holders of such liabilities. However, if the nonrestricted sector has securities which are good substitutes, then it will be more difficult for the restricted firms to shift the burden; deposits fall, both of which reduce the scale of their operations. Consequently the restrictions will be less effective in increasing the supply of mortgage funds. In terms of our analysis, the supply curve after restrictions shifts to the left towards its initial position, as other securities are substituted for the liabilities of the restricted institution.

Government guards against this outcome with a variety of complementary programs which either make it easier to shift the burden of restrictions to depositors by restricting the competition for funds, or by compensating the restricted institution with explicit and implicit sub-

sidies. For example, in the past Regulation Q has prevented commercial banks from freely competing with savings and loan institutions for time deposits. In the discussion which follows such policies are referred to as "cost shifting" policies.

Because portfolio restrictions can initially be implemented without cost to the government's budget, they have provided the illusion of being costless to society. In fact, the social cost of effective restrictions may be quite high and it has probably been borne largely by either the owners of the restricted institutions or the owners of particular kinds of deposits depending on the complementary policies adopted by government. With the recent implementation of explicit subsidies (to be discussed in Section II), portfolio restrictions now impose budget costs along with their less obvious other social costs.[5]

There are two ways that portfolio restriction programs may influence the effectiveness of wedge-type programs via an impact on the degree of substitutability between securities. First, going back to the perfect substitutability case depicted in Figure 2(a), if portfolio restrictions are sufficiently pervasive, the rightward shift in the supply curve of the restricted sector could result in the intersection of the demand curve with the supply curve of the restricted sector at a rate below i_{M_0}. The imposition of restrictions then reduces the degree of substitutability between securities for the market as a whole (from perfect substitutability to imperfect substitutability). The second way that portfolio restrictions may change the degree of substitutability between securities for the market as a whole is by changing it in an individual institution's portfolio. (See Silber.) For example, assume the regulation takes the form that the institution must hold at least 50 percent of its assets in mortgages. As the mortgage rate falls relative to other rates the institution can reduce its holdings to a minimum of 50 percent of its assets. After that the elasticity of substitution is zero and the elasticity of supply is zero (for a given level of deposits). Under both circumstances portfolio restrictions reduce the effectiveness of wedge-type programs by reducing substitutability between securities.

We now turn to the impact of mortgage characteristic programs on the supply of mortgage funds. A change in the characteristics of mortgages via, say, government guarantees or the establishment of a secondary market increases lenders' willingness to hold mortgages as investments. With such programs there is an impact on the volume of mortgage funds supplied even with a perfectly elastic mortgage supply curve. This occurs because the reduction in the risk of mortgages relative to the risk of other securities causes a downward shift even in the infinitely elastic supply curve in Figure 2(a).

An important side effect of the programs which lower the riskiness of mortgages is that they also make mortgages better substitutes for other securities. This follows as long as mortgages have greater default risk and a poorer secondary market compared with other securities. Under such conditions government programs make these risk characteristics of mortgages more like those of other securities, thereby increasing the degree of substitutability between mortgages and other securities.[6]

This side effect of mortgage characteristic programs implies that they increase the effectiveness of wedge-type programs and decrease the effectiveness of portfolio restrictions. This is a key interrelationship between existing housing programs as will be seen below. Note also that since the

[5] More extensive commentary on the social and budget costs of these programs appears in Penner and Silber.

[6] See Tobin for a discussion of this relationship between risk characteristics of securities and the degree of substitutability between securities.

mortgage characteristic programs change the slope of the supply curve as well as shifting it, one cannot make any simple statements about the relationship between the efficacy of these programs and the initial degree of substitutability.

II. An Application to Current Credit Programs

This section of the paper applies the theoretical outline set forth above to the current set of federal housing credit programs in the United States that are aimed at the private sector of the economy (i.e., excluding public housing). Existing housing credit programs are categorized according to the classifications outlined in the previous section. Specific analysis of the interaction between the various programs is also presented. It should be noted at the outset that our concern here is not with the income redistribution aspects of certain programs; rather, we are concerned solely with the impact of the program on mortgage credit and mortgage rates.

Table 1 sets forth the main characteristics of the existing housing programs. There are three groups of policies as outlined in the previous section: 1) wedge programs; 2) programs which operate on the risk characteristics of mortgages; and 3) portfolio restriction policies. In the first two columns of Table 1 a brief description of each program is presented as well as a listing of the agency responsible for its execution and the approximate size of the program during the fiscal year 1970.

The programs listed in the first 5 rows of the table are in the wedge-type category. Extensive comments are limited to those programs which have dual listings or may be ambiguous. The programs under sections 235, 236, and 243 (lines 1 and 2) are subsidies rationed out to particular classes of borrowers. Line 3 is the Housing Opportunity Allowance Program (*HOAP*) of the Federal Home Loan Bank System (*FHLBS*) provided under Title I of the Emergency Home Finance Act of 1970. A direct subsidy to mortgage borrowers is offered. Line 4 of Table 1 is the Tandem plan, implemented by the Government National Mortgage Association (*GNMA*) in "tandem" with the Federal National Mortgage Association (*FNMA*). The Tandem plan operates to produce a wedge when the FHA-ceiling rate is below market rates. Tax benefits (in particular, deductibility of mortgage interest in calculating taxable income) in line 5 also inserts a wedge between borrowers and most lenders. This wedge is operative as long as the income tax rate applicable to borrowers exceeds the tax rate paid by lenders (which is the likely case due to the favorable tax treatment of the major mortgage lenders, *S&L*'s, and mutual savings banks).

The next four lines of Table 1 list the housing credit programs aimed at altering the risk characteristics of mortgages. Line 6 records the secondary mortgage market operated by *FNMA* and Federal Home Loan Mortgage Corporation (*FHLMC*). The existence of these two institutions implies that mortgages are more liquid than otherwise because *FNMA* and *FHLMC* stand ready to make a market in existing mortgages.

This secondary market function of *FNMA* and *FHLMC* should be distinguished from the countercyclical and secular purchases of mortgages by these institutions which appear in line 11 of the table under portfolio restriction programs. *FNMA* in particular has behaved more like a financial institution restricted to mortgage holdings. Since both *FNMA* and the *FHLMC* have a government-imposed directive to support housing, they buy mortgages when there is a significant

TABLE 1—PROGRAM CHARACTERISTICS

Program	Agency and Program Size	Short Description	Source of Funds	Benefits as related to Substitutability Between: Mortgages and Other Primary Securities	Benefits as related to Substitutability Between: Deposits and Other Primary Securities
Wedge-Type					
1. Sec. 235 & 236	HUD In 1970, housing starts under these programs totalled 116 thousand units. The budget request for these two programs totalled $245 million. (Note all years refer to fiscal year.)	Subsidizes FHA-VA mortgage interest rate (to as low as 1 percent for low income single family (235) and multi-family (236) housing. Authority granted in Housing and Urban Development Act of 1968. (Programs were suspended Jan. 1973.)	U.S. Treasury	+	+
2. Sec. 243	HUD This program has never been funded. While not yet operational the administration has considered implementing the program.	Subsidizes FHA-VA or conventional mortgage interest rate (to as low as 7 percent) for middle income borrowers. Authority granted in Emergency Home Finance Act of 1970.	U.S. Treasury	+	+
3. Title I, *HOAP* (Housing Opportunity Allowance Program)	*FHLB* In 1971, $51 million was obligated. (Program no longer operative.)	*FHLB* subsidizes mortgage rate *S&Ls* charge their borrowers. As currently operated by the *FHLB*, a borrower's monthly payments are reduced by as much as $20 for a period of 5 years.	U.S. Treasury	+	+
4. Tandem	HUD, *GNMA*, *FNMA* The program was expanded in 1972. There is an authorization for $2.0 billion under the program.	When the effective rate on FHA-VA mortgages is above the ceiling rate, *GNMA* buys mortgages at par and resells them at a discount to *FNMA*. This insures the borrower of not paying the discount points in the form of, say, higher prices for a given house. Developed under HUD Act of 1968.	U.S. Treasury	+	+
5. Tax Benefits	U.S. Treasury, Congress	Homeowners may deduct mortgage interest payments in calculating taxable income for the personal income tax.	U.S. Treasury	+	+
Mortgage Characteristics					
6. *FNMA/FHLMC* Secondary Mortgage Market	*FNMA*, *FHLMC*	*FNMA* and *FHLMC* stand ready to buy and sell existing mortgages at market prices. *FNMA* was established in 1934 under the National Housing Act of 1934. The status of *FNMA* was altered in the Housing Act of 1954 and the HUD Act of 1968. The *FHLMC* was created by the Emergency Home Finance Act of 1970.	Borrowing from the public via agency issues plus back-up borrowing capability at the Treasury.	0	0

TABLE 1—(*Continued*)

Program	Agency and Program Size	Short Description	Source of Funds	Benefits as Related to Substitutability Between: Mortgages and Other Primary Securities	Benefits as Related to Substitutability Between: Deposits and Other Primary Securities
7. VA subsidized guarantees	Veterans Administration. During 1970 there was a net increase of $1.1 billion in guranteed mortgages.	Mortgages are guaranteed by the Veterans Administration for qualified veterans of the United States.	U.S. Treasury	0	0
8. FHA Unsubsidized insurance	HUD, FHA During 1970 there was a net increase of $4.6 billion in insured mortgages.	Mortgages can be insured by HUD under sections 203 and 207 of the National Housing Act of 1949. Premiums are paid by borrowers for the insurance.	Insurance premiums from borrowers. Back-up funds from Treasury equalled $339 million in 1970. Reserves from premiums paid totalled $1,203 million in 1970.	0	0
9. *GNMA* Guarantees of mortgage backed bonds.	*GNMA* During 1970 GNMA guaranteed $441 million of securities.	*GNMA* guarantees securities which are issued by private investors when their securities are backed by FHA-VA mortgages. Authorized under the HUD Act of 1968.	U.S. Treasury	0	0
Portfolio Restrictions					
a. Direct Measures					
10. *S&L* portfolio restrictions	*FHLB*	Savings and Loan Associations are permitted to hold only government bonds, mortgages and the issues of U.S. government agencies.		—	—
11. *FNMA* and *FHLMC* purchases	*FNMA* *FHLMC* During 1970 the net change in outstanding loans totalled $5.4 billion for *FNMA*. The *FHLMC* was not operational until mid-1971.	*FNMA* and *FHLMC* buy and sell mortgages in order to moderate cyclical swings in mortgage rates and credit.	Borrowing from the public via agency securities plus back-up borrowing capability at the Treasury (back-up funds equal $2.25 billion).	—	—
b. Cost-shifting Measures					
12. Deposit Insurance	*FHLB* via Federal Savings and Loan Insurance Corporation.	Deposits at *S&L*s are insured.	Premiums paid by *S&L*s plus back-up funds from Treasury. In 1970 paid-in reserves totalled $2.4 billion while the back-up authority was $750 million.	0	—

TABLE 1—(*Continued*)

Program	Agency and Program Size	Short Description	Source of Funds	Benefits as Related to Substitutability Between: Mortgages and Other Primary Securities	Benefits as Related to Substitutability Between: Deposits and Other Primary Securities
13. Regulation Q	Federal Reserve Board	Commercial bank time deposit rates are regulated by the Federal Reserve. Until 1966 such limitations curbed the competitive position of banks for savings. Since 1966 *S&Ls* have also had regulations on their deposit rates.	Not applicable.	0	—
14. *FHLB* unsubsidized lending	*FHLB* During 1970 the net change in outstanding loans totalled \$3.8 billion.	*S&Ls* are permitted to borrow at the *FHLB*. The rate charged depends upon the rate paid by the *FHLB* for funds.	Borrowing from the public via agency issues plus back-up borrowing capability equal to \$4.0 billion at the Treasury.	—	—
15. *FHLB* subsidized lending	*FHLB* Not implemented as yet.	The *FHLB* lends to *S&Ls* at subsidized rates. Authorized by Title I of the Emergency Home Finance Act of 1970.	U.S. Treasury	—	—
16. Favorable tax treatment of *S&Ls*	U.S. Treasury, Congress		U.S. Treasury	—	—

contraction in the volume of credit supplied by private institutions. Conceptually, if not empirically, these are quite distinct. Note that both *FNMA* and the *FHLMC* are subsidized by the government because each has a line of credit at the Treasury. In other words, the costs of the restrictions on *FNMA* and the *FHLMC* are shifted to the tax paying public because of the potential cost of the line of credit at the Treasury.

Line 7 lists the Veterans Administration mortgage guarantee program which insures holders of VA mortgages against default. The elimination of default risk increases the supply of such mortgage funds, lowering the contract mortgage rate and easing terms of credit. Since the borrowers do not pay the insurance premium the effective rate to the borrower is in fact lowered.

Line 8 lists the FHA-unsubsidized insurance program. Under this program the mortgage borrower pays an insurance premium to the government. While the contract mortgage rate is reduced (and terms of credit eased) after the guarantee is in existence, the effective rate (i.e., the

contract rate plus the insurance premium) to the borrower might very well be unchanged if the insurance premium were equal to what the private sector would charge (equal to expected loss plus administrative costs). There is one aspect of the program, however, which does produce a subsidy in the form of lower insurance premiums than otherwise, namely, the back-up funds authorized by Congress to be used in case defaults exhaust paid-in funds.

Line 9 lists another type of guarantee program by *GNMA*. Mortgage lenders who hold FHA-VA mortgages can now issue bonds insured by *GNMA* to finance these acquisitions. It is included under programs operating on mortgage characteristics because it has the same effect as increasing the "pledgeability" of mortgages as security.

There is an element of cost shifting in all of these programs operating on mortgage characteristics. By reducing the risks on mortgages they also counteract the burdens imposed by forcing an institution to hold mortgages. However, these programs are aimed primarily at voluntarily increasing the supply of mortgage funds from both restricted and unrestricted investors by lowering the risks on mortgages. The fact that they ease the burden imposed by restrictions just happens to be a desirable side effect.

The next group of programs comes under the heading of portfolio restriction policies. These policies are divided into two groups: (a) direct measures and (b) cost shifting measures. Lines 10 and 11 include the direct measures, i.e., portfolio restrictions on *S&L*'s (limiting them to holding mortgages, governments and agency issues) and the countercyclical and secular mortgage purchases of *FNMA* and the *FHLMC*.

The first two cost shifting programs are deposit insurance in line 12 and Regulation Q in line 13. Both of these measures try to offset the potentially adverse impact on deposits of the poorer risk/return tradeoff in *S&L* portfolios due to mortgage specialization. Note that in the case of deposit insurance, although premiums are paid there are back-up budget funds which presumably make premiums less than what would be charged privately. The subsidy in this case is similar to FHA-unsubsidized insurance.

Lines 14 and 15 include Federal Home Loan Bank (*FHLB*) subsidized and unsubsidized lending to *S&L*'s as cost shifting measures. The unsubsidized lending by *FHLB* contains some subsidy, hence, cost shifting, because the rate the *FHLB* pays on its liabilities is less than that of a private corporation because there is back-up authority for borrowing at the Treasury.

Line 16 lists the favorable tax treatment received by *S&L*'s as a cost shifting measure. While the tax treatment of *S&L*'s was made more stringent in the Tax Reform Act of 1969 they still retain significant tax benefits compared with unrestricted financial institutions.

A. *Interrelationships between and Impacts of Programs*

The first column of Table 1 lists the agency charged with administering the program and also provides a rough estimate of the size of the program during the fiscal year 1970. It is important to recognize that these numbers are not comparable between different programs. In some cases, such as 235 and 236, the subsidy as measured by actual budget costs is listed, while for others, such as *FNMA-FHLMC* operations, the total purchases of mortgages is listed. The implicit subsidy in the latter case is simply not known.

The ultimate measure of a program's effectiveness is its impact on mortgage

credit and housing production. While the table includes (where available) the number of houses constructed or the number of mortgage loans made under various programs, it is incorrect to use such numbers as measures of a program's impact. Clearly they do not record the *increment* in housing or credit due to the subsidy.

We noted in the beginning of Section I that increased mortgage credit tends to increase actual housing production. This is not necessarily the case. In fact the size of the impact may depend upon the nature of the particular credit program. For example, the funding of programs 235 and 236 does not permit the granting of subsidies to all who qualify legally. The funds must be rationed by administrative action. A key to the size of the impact of such credit on housing production is, therefore, the skill of the administrators in isolating potential homebuyers who are at the margin, i.e., who would withdraw from the market without the subsidy.

The specific nature of the subsidy calculation poses a related problem. The homeownership assistance subsidy under 235, for example, is set to either lower the mortgage interest rate to one percent or to lower the payment for the sum of mortgage interest, taxes and insurance to 20 percent of the owners income, whichever is less. If the size of the subsidy depends on the recipient's income (the second option) and he expects that his future income will not rise sufficiently to end the subsidy, then he has a strong incentive to maximize the size of the mortgage and to purchase the highest quality home since such quality improvements impose no personal marginal costs (they are borne entirely by the subsidy). This feature of the 235 program, i.e., relating the subsidy to the recipients' income, has probably raised the expenditures per housing unit. With a given budget for the 235 program, this might also imply a reduction of the quantity of new units constructed.[7]

In column 3 of Table 1 the source of funds for each program is listed. In some cases there is a primary source and a secondary or back-up source. As we noted earlier, the use of borrowing from the private capital market produces less of an impact than the use of general tax revenues. Similarly, the use of fees and charges produces a smaller impact than using general tax revenues. Obviously, however, direct costs are smaller in the borrowing and user-charges cases compared with tax revenues.

Columns 4 and 5 lists the relationship between types of substitutability and the benefits of each program. Benefits are defined as increased mortgage credit (or lower mortgage rates to borrowers). The question asked was as follows: for a given dollar level of activity of the program would mortgage credit increase (+), decrease (−), or remain the same (0) if there was an increase in the degree of substitutability between X and Y (where X and Y were either mortgages, other primary securities, or deposits of mortgage lending institutes). Note that the heading of the column could have been listed as "costs as related . . . " and the question asked could have been: for a given volume of mortgage credit (or rate charged to borrowers) would the dollar cost of the program increase, decrease, etc. . . . In other words, benefits and costs as defined here are related symmetrically to the substitutability question.

In Table 2 the interrelationships between the various programs are outlined. A posi-

[7] The redistribution goals of 235 and 236 programs may be better served by income-related subsidies. Yet even this is not clear. In the fourth quarter of 1971 the median annual income of families receiving 235 subsidies was $6,309, with only 3.9 percent of the families having an income below $4,000 (see p. 29 in *The Fourth Annual Report on National Housing Goals*).

TABLE 2—INTERRELATIONSIP BETWEEN PROGRAMS

On	Impact of Program															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1		0	0	0	0	+	+	+	+	0	0	na	na	na	na	na[a]
2	0		0	0	0	+	+	+	+	0	0	na	na	na	na	na
3	0	0		0	0	+	+	+	+	0	0	na	na	na	na	na
4	0	0	0		0	+	+	+	+	0	0	na	na	na	na	na
5	0	0	0	0		+	+	+	+	0	0	na	na	na	na	na
6	0	0	0	0	0		0	0	0	0	0	na	na	na	na	na
7	0	0	0	0	0	0		0	0	0	0	na	na	na	na	na
8	0	0	0	0	0	0	0		0	0	0	na	na	na	na	na
9	0	0	0	0	0	0	0	0		0	0	na	na	na	na	na
10	0	0	0	0	0	−	−	−	−		0	+	+	+	+	+
11	0	0	0	0	0	−	−	−	−	0		0	+	0	0	0
12	0	0	0	0	0	+	+	+	+	0	0		0	0	0	0
13	0	0	0	0	0	+	+	+	+	0	0	0		0	0	0
14	0	0	0	0	0	+	+	+	+	0	0	0	0		0	0
15	0	0	0	0	0	+	+	+	+	0	0	0	0	0		0
16	0	0	0	0	0	+	+	+	+	0	0	0	0	0	0	

[a] na = not applicable.

Key:
1. Sec. 235, 236
2. Sec. 243
3. Title I (*HOAP*)
4. Tandem
5. Tax benefits to public
6. *FNMA/FHLMC* secondary market
7. VA subsidized guarantees
8. FHA unsubidized insurance
9. *GNMA* guarantees of mortgage backed bonds
10. *S&L* portfolio restrictions
11. *FNMA/FHLMC* purchases
12. Deposit insurance
13. Regulation Q
14. *FHLB* unsubsidized lending
15. *FHLB* subsidized lending
16. Favorable tax treatment of *S&Ls*

tive relationship (+) between two programs occurs when the initiation, or increased level of activity, of program X causes an increase in the benefits or a reduction in the costs (as defined above) derived from a given level of program Y. Zero and negative relationships are defined in the same way with the appropriate changes.

Columns 1 through 5 indicate that wedge-type programs have no effect on the efficacy of any of the other housing programs. The impacts of the mortgage characteristic programs (6, 7, 8, and 9) are based on the assumption that, in general, mortgages have greater default risk and marketability risk compared with other securities (for example, corporates and governments) and that these risk differences imply a lower degree of substitutability between mortgages and other securities than would otherwise exist. Any program which reduces the default and marketability risks of mortgages (making these risks more like those on other securities) also makes mortgages better substitutes for other securities. The positive relationship between mortgage-characteristic programs and the wedge-type programs (1–5) and the negative relationship between mortgage characteristic programs and portfolio restrictions (10–11) follows directly. Since the five cost-shifting programs (12–16) seek to offset the costs of restrictions, the effectiveness of any given cost-shifting measure is increased the lower is the "cost" of the restrictions, i.e. the greater the similarities between mortgages and other securities. Hence, there is a positive relationship between the mortgage characteristic programs and the cost-shifting measures.

The relationship between the mortgage characteristic programs and the other programs would be reversed if the marketability and default risks of mortgages were lower than those on other securities. This situation cannot occur as long as there is a security such as government bonds which have zero default risk and the "most efficient" secondary market possible given the existing financial technology.

The impact of any one of the mortgage characteristic programs on any of the others is probably zero unless one wishes to argue that there are "diminishing returns" to guarantees on top of guarantees. Under such circumstances, the existence of FHA and VA guarantees is negatively related to the effectiveness of *GNMA* guarantees of mortgage-backed securities.

The last group of programs, portfolio restriction policies, is a self-contained group as far as interrelationships are concerned. Neither *S&L* portfolio restrictions nor countercyclical mortgage purchases by *FNMA* and the *FHLMC* have any impact on the efficacy of other programs.[8] In a very limited sense the existence of restrictions on *S&L*'s makes the cost-shifting measures relevant; however, an increased level of restrictions does not make a given level of cost shifting more effective.

The impacts of the cost-shifting measures on *S&L* restrictions is clearly positive since, if these cost-shifting programs did not exist, the risk/return position of the restricted industry would be worsened and the captive demand for mortgages would be reduced (as firms left the industry and as depositors substituted other securities for deposits). The only cost-shifting measure which increases the efficacy of *FNMA* and *FHLMC* mortgage purchases is Regulation Q (since it reduces the competition for *FNMA-FHLMC* obligations from commercial bank time deposit liabilities).

[8] We noted above that portfolio restrictions on *S&L*'s might reduce the degree of substitutability between mortgages and other securities. Since there is uncertainty with regard to this effect we chose not to include in the table a negative impact of portfolio restriction policies on the wedge-type programs.

III. Conclusions

The key determinant of the effectiveness of two of the three housing finance programs is the degree of substitutability between mortgages and other categories of securities. Wedge-type programs are most effective with high substitutability while portfolio restrictions are most effective with poor substitutability. In a world of perfect certainty the degree of substitutability would be known, hence the precise benefits (in terms of mortgage credit diverted or reduction in rates) could be calculated for each program. These benefits could be compared with the associated costs and an unambiguous optimal policy to divert funds to the mortgage market would emerge.

Policy decisions, however, must be made under conditions of uncertainty with regard to the parameters which determine the effectiveness of each policy. Under such circumstances it has been shown that diversification of policy tools is the optimum strategy. (See William Brainard.) Hence, the simultaneous use of wedge-type programs and portfolio restriction policies is a desirable strategy.

The fact that the effectiveness of mortgage characteristic programs is independent of substitutability between mortgages and other securities suggests that they too should be included in the diversified policy strategy. It does not imply that exclusive reliance be placed on such policies since the cost per increased output of houses may be much larger than for either of the other programs.

The fact that mortgage characteristic programs increase the degree of substitutability between mortgages and other securities has one clear implication: wedge-

type programs become more effective while portfolio-restriction policies become less effective. This might help justify or explain the relatively late emergence of wedge-type programs compared with portfolio restrictions. As mortgage guarantees and the establishment of a secondary market made mortgages better substitutes for other securities, the effectiveness of portfolio restrictions declined and the effectiveness of wedge-type programs increased. However, not only do portfolio restrictions become less effective ways of increasing mortgage funds, they also become less harmful to the restricted institution. The reduction in the burden of restrictions caused by the expansion of the programs which reduce risk (as well as the provision of subsidized loans to $S\&L$'s) has probably been important in allowing government to suspend part of Regulation Q and to recommend its permanent elimination.

REFERENCES

E. A. Brady, "An Econometric Analysis of the U.S. Residential Housing Market," working paper no. 11, Federal Home Loan Bank System, Washington 1970.

W. Brainard, "Uncertainty and the Effectiveness of Policy," *Amer. Econ. Rev. Proc.*, May 1967, *57*, 411–26.

G. F. Break, "Federal Loan Insurance for Housing," *Federal Credit Agencies: Commission on Money and Credit*, Englewood Cliffs 1963.

R. C. Fair, *A Short-Run Forecasting Model of the United States*, Lexington 1971.

J. Guttentag, "The Federal National Mortgage Association," *Federal Credit Agencies: Commission on Money and Credit*, Englewood Cliffs 1963.

R. G. Penner and W. L. Silber, "Federal Housing Credit Programs: Costs Benefits and Interactions," part 5, *The Economics of Federal Subsidy Programs*, Joint Economic Committee, Washington 1972.

D. C. Rao and I. Kaminow, "Selective Credit Controls and the Real Investment Mix: A General Equilibrium Approach," *J. Finance*, forthcoming.

W. L. Silber, "Portfolio Substitutability Regulations and Monetary Policy," *Quart. J. Econ.*, May 1969, *83*, 197–219.

C. Swan, "Homebuilding: A Review of Experience," *Brookings Papers*, Washington 1970, *1*.

J. Tobin, "An Essay on Principles of Debt Management," *Fiscal and Debt Management Policies: Commission on Money and Credit*, Englewood Cliffs 1963.

U.S. Congress, *The Fourth Annual Report on National Housing Goals*, House Document No. 92–319, Washington 1972.

On Monopoly Welfare Losses

By ABRAM BERGSON*

Are welfare losses due to monopolistic pricing apt to be consequential in an advanced economy such as that of the United States? Generations of economists have assumed that they are, but in 1954 Arnold Harberger argued that in the United States during the late 1920's, the total of such losses in manufacturing was equivalent to but "... a tenth of a percent of national income" (p. 87). Subsequently, David Schwartzman found that in the United States in 1954 welfare losses due to monopolistic pricing in industry were equivalent to "... less than $234 million, or less than 0.1 percent of the national income" (pp. 629–30). Still more recently Harvey Leibenstein, relying partly on the findings of Harberger and Schwartzman, has also argued that departures from competitive norms such as result under monopolistic pricing are likely to be relatively inconsequential. Generally, "... they hardly seem worth worrying about" (Leibenstein, p. 395).

From such findings, George Stigler was an early dissenter and further calculations of David Kamerschen indicated that welfare losses due to monopoly pricing may be appreciably greater than Harberger and Schwartzman computed. The contention that such losses are inconsequential nevertheless has apparently gained no little currency. Perhaps that is as it should be, but the underlying methodology still does not seem to have been sufficiently explored. Some of the more basic empirical assumptions may also be more questionable than has been supposed. A brief reexamination stressing those aspects may facilitate appraisal of what is surely one of the more important current issues in economics.

I. Some Methodological Considerations

The evaluations of monopoly welfare losses by Harberger, Schwartzman, and Leibenstein all represent applications of consumer's surplus analysis. As such, they are in a sense rather special, for expressly or by implication, all these writers apply a well-known formula of Harold Hotelling for such surplus. Their calculations may also be viewed, however, as measurements of consumer's surplus without reference to the Hotelling formula. As a preliminary, I briefly reformulate the relevant principles, which still seem somewhat in need of clarification.

What counts for Harberger, Schwartzman, and Leibenstein is efficiency rather than equity. I assume, therefore, as is often done in similar contexts, that there is but one household in the community. The household, however, behaves competitively in purchasing consumer's goods. As will appear, the analyses being considered also assume constant costs. For present purposes it is more convenient to work with the somewhat more restrictive supposition that the community's production possibilities are linear.

As J. R. Hicks taught us long ago, consumer's surplus is susceptible to diverse constructions. The particular construction need not be a practically very important matter, but we may conveniently consider the evaluations that have been made in relation to a concept of surplus corresponding to the compensating variation as understood by Hicks (see pp. 61ff, 69ff),

* Harvard University.

i.e., the compensatory change in income needed to assure that a household's utility is unaffected by a change in price.

For Hicks, however, the compensating variation represents a change from an otherwise given income. Such a variation is properly taken to measure welfare losses or gains if income is indeed constant. Our household, however, is subject not only to changes in prices, those being the changes due to the fixing of the prices in monopolistic industries at monopolistic rather than competitive levels, but also to a concomitant variation in income. In the analysis of monopoly welfare losses, the concern is with the impact of misallocation rather than unemployment. Hence, with the prices of monopolistic products at monpolistic rather than competitive levels, resources released from monopolistic industries supposedly are reemployed in competitive ones. That is possible only if household income is higher when prices in the monopolistic industries are at monopolistic levels than when they are at competitive levels. Indeed, income must increase by the amount of monopoly profits, for only in that way will there still be sufficient income to employ all the primary factors previously employed.

In short, we are confronted with variations within a general equilibrium context. In consumer's surplus analysis the usual practice even so is to focus on partial equilibrium, but the analysis here seems facilitated if general equilibrium is put clearly to the fore, and Hicks' compensating variation is viewed accordingly.

Thus, a distinction is made between the *gross* compensating variation (GCV), representing the change in income needed to compensate households for some specified variation in prices, and the *net* compensating variation (NCV), representing the gross change, less the adjustment in income (MIA) required to maintain full employment in the face of such a variation in prices, i.e.,

$$NCV = GCV - MIA \tag{1}$$

The GCV corresponds to what Hicks called compensating variation, but the NCV is of primary interest here.[1]

Of particular concern now is a rise in prices in monopolistic industries to monopolistic from competitive levels. Referring to such price changes, it suffices here if there is but a single competitive industry in the community considered. If for the moment we also assume that there is only one monopolistic industry, the price and income variations of interest may be illustrated graphically. In Figure 1, x_1 is the product of the monopolistic and x_2 the

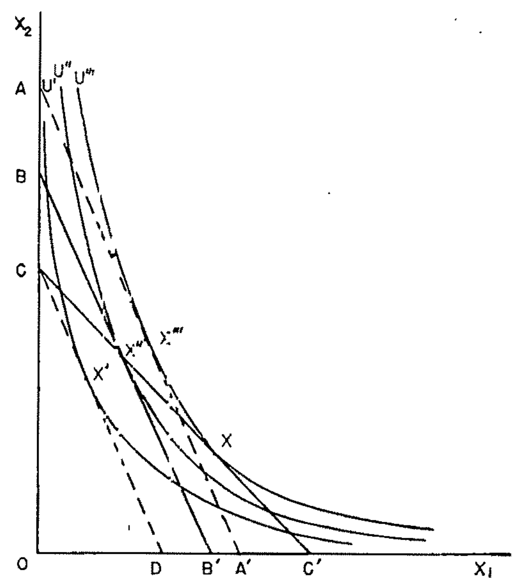


FIGURE 1

[1] For simplicity I assume throughout that monopoly price-cost ratios are unchanged under the impact of the changes in household income in question, but it is easy to see that no essentials would be affected if the ratios should vary. What counts are the monopolistic price-cost ratios when income is again at the full employment level, and the GCV could simply be understood as the addition to the household's initial income level that is needed to compensate for the charging of those price-cost ratios. The MIA is calculated as before as the difference between the two full employment income levels, one where monopolistic products are priced monopolistically and the other where they are priced competitively.

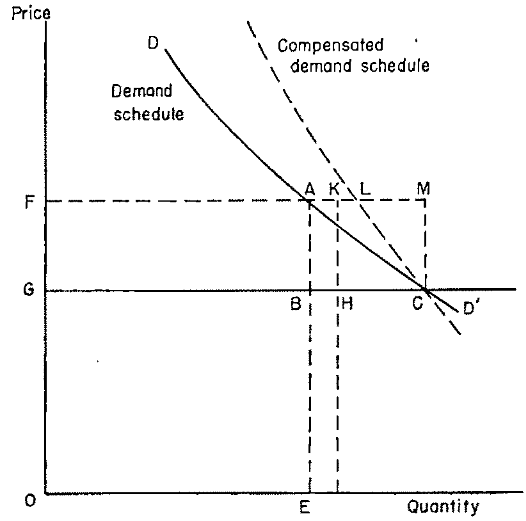
product of the competitive industry, and CC' represents, the community's production possibilities. Several of the household's indifference curves are also shown, and the point x on U''' represents the household's consumption as it would be if there were competitive pricing throughout the economy, i.e., price equals marginal costs in both the monopolistic and the competitive industries. At x' on U' the household's consumption is as it would be if under monopolistic pricing for x_1 the household's income (with x_2 as the numeraire) should be constant at the competitive level OC. With consumption at x', however, there would be unemployed resources, and that is avoided only if income rises by BC to OB, permitting consumption under monopolistic pricing for x_1 to be at x''. That point is again on the production possibilities schedule, and also on U''. The household has still suffered a welfare loss, though, as a result of monopolistic pricing for x_1. The loss is represented by the shift from U''' to U'', but this would be fully offset were there a further increase in income by AB, that would permit consumption to rise to x''' on U'''. Such consumption, of course, is not feasible, but the increase in income by AB still can serve as a hypothetical measure of the welfare loss due to monopoly.

Thus, AB represents the NCV for monopolistic pricing as that has been defined. We also have

$$AB = AC - BC \tag{2}$$

where AC corresponds to the GCV for monopolistic pricing, and BC represents the MIA, or increase in income above the competitive level that is needed to maintain full employment under monopolistic pricing for x_1.

I have delineated the NCV in terms of the household's indifference map. Could it be calculated from a conventional Marshallian demand schedule for the monopolized product? Subject to one proviso, it could be, for the NCV then precisely corresponds to the triangle-like area ABC in Figure 2. Here DD' represents the Marshallian demand schedule for x_1 that prevails when household income is at a level that would assure full employment under competitive pricing for x_1. Since costs are constant, such pricing means that x_1 would sell at average cost, equalling OG. Under monopoly, however, the price is increased to OF.



OF = Monopoly Price
OG = Competitive Price
FK = Monopoly Output After Adjustment in Money Income to Assure Full Employment

Figure 2

The proviso is the proverbial one that the income elasticity of demand for the product in question be zero. In that case, the budget points x', x'', and x''' in Figure 1 all include the same x_1 and differ only in respect of x_2. Also, the GCV, or AC in that figure, corresponds to the area ABC together with the rectangle $FABG$ in Figure, 2, for "income effects" that might distort the relation between the sum of those areas and the GCV are zero (see Hicks, pp. 68ff). It also follows that the MIA, or BC in Figure 1, corresponds to the rec-

tangle *FABG* in Figure 2. Given that the consumption of x_1 in x' is the same as in x'', the x_1 in x'' corresponds to *FA*. Hence, *FABG* represents monopoly profits at a full employment income level, and also the addition that must be made to income if the resources released from x_1 by the introduction of monopolistic pricing there are to be fully absorbed into x_2. It also follows that the *NCV*, or *AB* in Figure 1, corresponds to the area *ABC* in Figure 2, which is in effect the difference between the *GCV* and *MIA*.

What if the income elasticity of demand is other than zero? As is well known, calculation of the Hicksian compensating variation from a Marshallian demand schedule in that case is subject to error. That must be so also, therefore, for the *GCV*, but of interest here is the *NCV* so we must also consider that the *MIA*, as just calculated, is likewise subject to error. Curiously, the errors tend to be offsetting.

Thus, suppose the monopolized product is a superior one. Then the *GCV* for an increase in the price of that product from the competitive to the monopolistic level now corresponds to the area that is bounded by that price change to the left of the Hicksian or compensated demand schedule, that is, the area *FLCG* in Figure 2. That area exceeds the corresponding area, *FACG*, to the left of the Marshallian or uncompensated demand schedule, which previously represented the *GCV*. On the other hand, monopoly profits at a full employment income level and hence the *MIA* also increase, and are now represented by the area *FKHG* in the figure, rather than the area formerly considered, *FABG*. That follows at once from the fact that *FK* corresponds to x_1 in x'', and is thus intermediate between x_1 in x', or *FA*, and x_1 in x''', or *FL*.

Since the monopolized product must be supposed to be absorbing a large part of the household's income, the first of these income effects should often be rather large. Presumably that would ordinarily be true also of the second. But as indicated the two effects should also be mutually offsetting. How the *NCV*, now given by the area *KLCH* (i.e., the excess of *FLCG* over *FKHG*), compares with the area *ABC*, is nevertheless not an easy subject for generalization. For a superior good, however, it might be supposed that, as given by the *NCV*, the welfare loss would, if anything, be understated by an area such as *ABC* in Figure 2. Curiously that need not be so, for the error in the *GCV* might be more than offset by the further one in the *MIA*. For diverse linear demand schedules it can be shown that that is in fact the case.[2]

Proceeding more formally, we may readily extend the analysis to many monopolized products. Suppose there are n goods that are monopolized, and that a single product, $(n+1)$, is competitive. If we take that as numeraire, there are n prices, $p_1 \ldots p_n$, which vary from $p_1^c \ldots p_n^c$ to $p_1^m \ldots p_n^m$. Household income is I, and in the initial equilibrium,

[2] Reference is to demand schedules for the monopolized product of the form:

$$x = ap + bI + c$$

where x is the consumption of the monopolized product (for convenience, I omit subscripts at this point), p is its price, I is income, and a, b, and c are constants. This formula represents a family of Marshallian or uncompensated demand schedules, each such schedule corresponding to a given level of I. From the Slutsky equation, I derive a formula for the corresponding Hicksian or compensated demand schedules, and on this basis calculate the value of *NCV* corresponding to the area *ABC* and that corresponding to the area *KLCH* in Figure 2 under alternative assumptions as to the uncompensated price elasticity (η_{xp}) and income elasticity (η_{xI}) of demand in the vicinity of the full employment equilibrium with monopolistic pricing for x. With $\eta_{xp}=1.5$ and $\eta_{xI}=1.0$, I find that the *NCV* corresponding to *ABC* is 1.04 and that corresponding to *KLCH*, .77 percent of household income. The corresponding measures of *NCV* when $\eta_{xp}=6.0$ and $\eta_{xI}=3$ are 4.17 and 3.88 percent of household income. I assume that x accounts for one-half the household budget under full employment equilibrium with monopolistic pricing for x, and that with that pricing p is 20 percent above costs.

I^c. Then the household's utility may be written as a function:

$$(3) \qquad U = U(p_1, \ldots, p_n, I)$$

Hence, the GCV is such that

$$(4) \quad U(\overset{c}{p}_1 \ldots \overset{c}{p}_n, I^c) = U(\overset{m}{p}_1 \ldots \overset{m}{p}_n, I^c + GCV)$$

According to a well-known formula due mainly to Hicks (pp. 169ff), we also have

$$(5) \qquad GCV = \sum_{i=1}^{n} gcv_i$$

In (5), GCV is as before, though now changes in many prices together are being compensated for. Also, gcv_i is the gross compensating variation for the increase in the price of the ith of n monopolized products. The gcv_i for those products, however, must be determined sequentially. Thus, if they are treated in the order in which the products are numbered, then for gcv_j reference is to the income increment that would be needed to compensate for the increase in the price of the jth product when prices for all products $l \ldots (j-1)$ have already changed and have been compensated for.[3]

[3] So far as it is understood that in the determination of gcv_j, previous price changes have been compensated for, this formulation seems somewhat novel. That gvc_j must be so construed, however, is seen at once when we consider that with that construction:

$$U(\overset{c}{p}_1, \ldots, \overset{c}{p}_j, \ldots, \overset{c}{p}_n, I^c) = U(\overset{m}{p}_1, \overset{c}{p}_2, \ldots, \overset{c}{p}_j, \ldots, \overset{c}{p}_n, I^c + gcv_1) = \ldots$$

$$U\left(\overset{m}{p}_1, \ldots, \overset{m}{p}_{j-1}, \overset{c}{p}_j, \ldots, \overset{c}{p}_n, I^c + \sum_{i=1}^{j-1} gcv_i\right) = \ldots$$

$$= U\left(\overset{m}{p}_1, \ldots, \overset{m}{p}_n, I^c + \sum_{i=1}^{n} gcv_i\right)$$

Then,

$$U(\overset{m}{p}_1, \ldots, \overset{m}{p}_n, I^c + GCV) = U\left(\overset{m}{p}_1, \ldots, \overset{m}{p}_n, I^c + \sum_{i=1}^{n} gcv_i\right)$$

Since utility is supposed to vary monotonically with I, (5) follows.

It also follows that GCV may still be evaluated from uncompensated demand schedules, but for each gcv_i constituting it reference, strictly speaking, must be to that uncompensated schedule for x_i conforming to the indicated determination of gcv_i. Thus, suppose the demand function for the jth product is given by

$$(6) \qquad x_j = d^j(p_1 \ldots p_j \ldots p_n, I)$$

Then gcv_j is to be calculated from the particular member of that family given by

$$(7) \quad x_j = d^j\left(\overset{m}{p}_1 \ldots \overset{m}{p}_{j-1}, p_j, \overset{c}{p}_{j+1} \ldots \overset{c}{p}_n, I^c + \sum_{i=1}^{j-1} gcv_i\right)$$

Each and every gcv_i as so calculated is precisely correct and all together sum to GCV if income elasticities of demand for all monopolized products are zero, but otherwise calculated gcv_i and the resulting GCV will err in one way or another. Of course, if income elasticities were zero, there would also be no need to be concerned about the precise level at which money income is held constant in each demand schedule, for example, whether at I^c or at I^c plus a sum such as appears in (7).

In order to obtain the NCV, we must deduct from GCV the MIA, the adjustment in income needed to assure full employment in the face of all price changes together. Here

$$(8) \quad MIA = \sum_{j=1}^{n} mia_j = \sum_{j=1}^{n} (\overset{m}{p}_j - \overset{c}{p}_j)\overset{m}{x}_j$$

where $\overset{m}{x}_j$ is the consumption of x_j, when all prices of monopolized goods are at monopoly levels and income has been increased to $(I^c + MIA)$, i.e.,

$$(9) \quad \overset{m}{x}_j = d^j(\overset{m}{p}_1 \ldots \overset{m}{p}_j \ldots \overset{m}{p}_n, I^c + MIA)$$

MIA is determined in principle, along with x_j^m, from (8) together with (9) for $x_j^m, j=1 \ldots n$.

Might we also calculate *NCV* as so determined simply by summing areas such as *ABC* in Figure 2? To calculate such areas, reference here presumbaly should be made to demand schedules corresponding to that for x_j in (7) when all prices other than p_j are at competitive levels while I corresponds to I^c, or full employment income when prices are at competitive levels. On that understanding, the answer to the question posed is in the affirmative under two conditions: (i) The income elasticity of demand for every monopolized good is zero. (ii) Cross elasticities of demand among all monopolized products are zero.

Actually, under these conditions it is immaterial at what levels prices of products other than the one considered and income are held constant in the calculation of an area such as *ABC*, for then all terms other than the price of the product in question drop out of the formulas for the demand for x_j in (7) and (9). The uncompensated demand schedule for any good also coincides with its compensated demand schedule. Hence for any x_i, the areas *FACG* and *FABG* in Figure 2 precisely correspond to gvc_j and mia_j. The difference between the aggregate of areas such as *FACG* and the aggregate of rectangles such as *FABG*, and hence also the resulting aggregate of areas such as *ABC*, for all commodities, also precisely correspond to the *NCV*.

What if the stated conditions are not met? In that case, the aggregate of areas such as *FACG* can only be expected to diverge from the *GCV*, since for any x_j such an area will differ from the gcv_j. That will be so insofar as: (i) The gcv_j is properly calculated from a demand schedule such as that in (7), where prices for products other than x_j are held constant, some at monopolistic and some at competitive levels, and income is held constant at a level corresponding to that indicated in (7); and (ii) There are income effects of the usual sort, i.e., those manifest in a divergence between the compensated and uncompensated demand schedules for x_j. Similarly, the sum of areas such as *FABG* can hardly be expected to correspond with the *MIA* as given by (8) and (9). As in the case of a single monopolized product, errors at different points may sometimes tend to be offsetting. So far as products generally are more often substitutes than complements, that, as easily seen, is apt to be so regarding errors due, on the one hand, to the difference between uncompensated and compensated demand schedules and, on the other, to the difference between the sum of rectangles such as *FABG* and the *MIA* as given by (8) and (9). Nevertheless, the relation of the sum of areas such as *ABC* in Figure 2 to welfare losses as given by the *NCV* is rather complex. Without a detailed knowledge of demand functions, in all their dimensions, i.e., demand functions such as (7), we cannot really tell how nearly, if at all, the one is approached by the other.

II. The Case for Inconsequentiality

To come to Harberger, Schwartzman, and Leibenstein, their case for the inconsequentiality of welfare losses due to monopolistic pricing is too familiar to need any detailed replication here. Moreover, the chief question that I must raise about their methodology is already evident. Thus, Harberger and Schwartzman both calculate monopoly welfare losses simply by aggregating for different industries areas such as *ABC* in Figure 2. That is also Leibenstein's procedure, though beyond collating previous findings he limits himself to illustrative calculations. The procedure, to repeat, is also a valid one,

provided that the income elasticity of demand for every monopolized product and all cross elasticities of demand among such products are zero. Otherwise, however, the procedure is subject to error at diverse points; an error, moreover, which is difficult to gauge. The practical import of the Harberger, Schwartzman, and Leibenstein findings, therefore, is obscure.

But I have considered their methodology from a more or less conventional standpoint. Is their procedure not more defensible in terms of the Hotelling formula that they apply? That seems doubtful when we consider that Hotelling derives his formula simply by disregarding higher order terms in a Taylor expansion of the utility function. As he properly explains, the formula yields "an approximate measure when deviations from the optimum system are not great" (p. 253). While Harberger, Schwartzman, and Leibenstein consider deviations from the optimum represented by varying but often sizable divergencies of prices from costs, they apparently assume that the Hotelling formula still yields a good approximation. They could be right, but it is difficult to know that without further inquiry.

According to the Hotelling formula, furthermore, welfare losses resulting from a displacement from a competitive equilibrium, and hence optimum, amount in money terms to $\frac{1}{2}\sum dp_i dx_i$. Here dp_i and dx_i are the variations in p_i and x_i that the displacement entails. The summation is taken over all goods, though if, as here, one commodity is taken as numeraire, that necessarily drops out of the formula. Note, however, that dx_i is the change in the consumption of x_i that is induced by a change not only in the price of that good but in the prices of all other goods. Supposedly income also varies in a prescribed manner.[4] Hence complex demand effects of the sort I encountered above in the calculation of the NCV are also encountered here. As already indicated, such effects are nevertheless neglected by Harberger, Schwartzman, and Leibenstein.

Finally, Hotelling takes welfare losses to be represented in money terms by the quotient dU/α, where dU is the increment in utility and α is the marginal utility of money. For variations of any magnitude, however, welfare losses as so understood are unique only if the marginal utility of money is assumed constant. That was Marshall's assumption, but he would not have made it where prices are varying for products constituting a major part of the household's budget. Constancy in the marginal utility of money also entails a further assumption of cardinal measurability of utility, which Hotelling himself apparently preferred to avoid.[5]

[4] In his essay, Hotelling focuses on the impact of the introduction of excise in place of income taxes yielding the same total revenue. Since what counts is money income net of income taxes, the prescribed change in money income is evidently equal to the total revenue from the excise taxes, and so corresponds precisely to the MIA as defined here. As it turns out, that also means that the displacement from the optimum that Hotelling considers, like the one I consider, entails a shift along a plane in which consumption at predisplacement prices is constant. In my analysis, however, that occurs because of the assumption of a linear transformation locus. I find it rather puzzling, therefore, that Hotelling, pp. 255–56, apparently considers his formula as applying even where marginal costs are variable.

[5] Might we not, however, interpret α as representing simply the marginal utility of money at some intermediate point in the range in question, and so avoid in this way the assumption of constancy of the marginal utility of money? If we may judge from a recent discussion by Harberger (1971, pp. 786ff) of a formula similar to Hotelling's, he might now argue that α in the Hotelling formula be construed in just that way. But that apparently would mean that the term

$$(\alpha/2) \sum dp_i dx_i$$

which was derived as a lower order term in a Taylor's expansion, is nevertheless to be interpreted as the remainder term. How that may be done is not clear. Moreover, it is still disconcerting that the measure of welfare losses depends on an unknown point to which reference is made within the range in question. While Harberger applies the Hotelling formula without qualification, it should be observed that Schwartzman, p.

As seen by Harberger, Schwartzman, and Leibenstein, monopoly welfare losses evidently must turn primarily on two empirical factors: price elasticities of demand and price-cost ratios for monopolized products. That is still true according to the alternative methodology that I have elaborated, though Harberger, Schwartzman, and Leibenstein refer to own price elasticities, while as seen here cross and income elasticities also matter (and in the latter case, as seen from (7) and (9), not merely through the usual sort of income effects). It should be observed, therefore, that Harberger assumes throughout a price elasticity of unity for monopolized products. In his view,

> . . . one need only look at the list of industries . . . considered in order to get the feeling that the elasticities in question are probably quite low. The presumption of low elasticity is further strengthened by the fact that what we envisage is not the substitution of one industry's product against all other products, but rather the substitution of one great aggregate of products . . . for another aggregate. [1954, p. 79]

The assumption of a price elasticity of unity, however, was held "objectionable" by Stigler for the reason that "A monopolist does not operate where his marginal revenue is zero. . . . In any event, . . . most industries have long-run demand curves which are elastic" (p. 34). According to Schwartzman, Harberger's assumption is indeed to be thought of as applying to an industry rather than an individual firm: "Moreover, if we are interested in the value of resource misallocation by monopolistic industries as a group, the relevant demand elasticity is less than the average of the individual industry demand elasticities" (pp. 628–29). Schwartzman assumes the price elasticity of demand is no more than 2. For purposes of illustrative calculation, Leibenstein assumes that in monopolized industries the " . . . average elasticity of demand is 1.5" (pp. 395–96).

How elastic the demand for a composite of all monopolized commodities is in respect of price is an interesting question. But for purposes of appraising monopoly welfare losses, it is also not too relevant. All monopolized products could properly be treated here as a single composite product only if the prices of all such products exceed their competitive levels by relatively the same amount. So far as the facts are otherwise, welfare losses originate not only in inordinately high monopoly prices generally, but in the variation for different products in the relation of such prices to costs. Also, what counts is variation not only between industries in any conventional sense, but between products; even between different qualities or models of the same product (for example, different models of Buicks) so far as price-cost ratios vary here as well. In principle, therefore, welfare losses must be calculated by reference to the demand for each and every such monopolistic product, quality or model as the case may be. Only in that way can all such losses be allowed for.[6]

As to how elastic demand is for one or another product, quality or model, that

630, notes that he is neglecting variations in the demand for a product associated with changes in equilibrium magnitudes generally, as distinct from a change in the price of that product alone. He apparently considers such variations unimportant. Leibenstein does not refer to Hotelling, but he accepts Harberger's methodology. Leibenstein, p. 396, n. 3, refers, however, to the error due to neglect of income effects originating in the divergence of uncompensated from compensated demand schedules, and concludes that such effects are apt to be negligible. He may be right but as indicated income effects are seen here to be much more complex than he thus assumes.

[6] As explained, Harberger, Schwartzman, and Leibenstein all assume constant costs. In this essay, I have fallen in with that assumption, but note that so far as costs vary even uniform monopoly price-cost ratios would not mean that monopoly prices are all in the same relation to those that would prevail under competitive pricing.

must depend on the item. Presumably it also depends on the prices of other goods and household income. That is a matter of some importance, since in the calculation of gcv_i reference is to demand schedules which vary from item to item in respect of the levels at which prices of related goods and income are held constant. Any a priori assumption must be a gross oversimplification in this sphere, but I doubt that an elasticity of unity or even of 2 can be nearly large enough to encompass the interesting range of possibilities for the purposes of a calculation such as in question.

In this connection, it is pertinent to recall the relation familiar in monopoly theory that is alluded to by Stigler:

$$\overset{i}{\eta}_{xp} = p_i/(p_i - MC_i) \tag{10}$$

Here η^i_{xp} is the price elasticity of demand for the ith product, p_i is its price and MC_i, its marginal cost. No one supposes that (10) can be more than very broadly applicable to real world monoplies, but Leibenstein, for example, assumes a monopoly price-cost ratio of 1.2. It seems illuminating that with that ratio (10) would imply an elasticity of demand as high as 6. Leibenstein considers that the monopolistic price-cost ratios must in fact average less than 20 percent, but for a ratio of, say 10 percent, (10) yields an elasticity of 11. To repeat, though, (10) hardly applies exactly in practice, and it may be just as well to note here that I myself will allow later for possible divergencies from it in either direction. This assumption is in order when we consider such aspects as the limitations in a firm's knowledge of its cost and demand schedules; the fact that because of price leadership, collusion and other adaptations to oligopolistic interdependence the firm may calculate in terms of a demand schedule that is less elastic than the one relevant here (i.e., that where prices of all other products, including even close substitutes, are unchanged), and so on.

What of the price-cost ratios? Harberger deduced these from data on profit rates for different industries. Stigler, pp. 34–35, held, however, that monopoly profits are often capitalized, and in diverse ways not always sufficiently allowed for by Harberger, with the result that the indicated price-cost ratios may often be too low. Still other forces could operate to produce the same result. As Harberger is aware, monopoly prices reflect not only monopoly profits but monopolistic advertising, but, even for the early period he considers, the resulting additional misallocation cannot be dismissed as he dismisses it simply on the ground that advertising outlays were "... well under 2 percent of sales for all the industries ..." (1954, pp. 85–86) studied. The industries studied include competitive as well as monopolistic ones, so the corresponding ratio for monopolistic industries alone must be higher. What counts here, moreover, is not only average relations but their variation between industries and products.

Harberger treats intermediates as if they were consumers' goods. The relevant monopoly profits for any consumer's good, however, are not those prevailing in the industry in question, but those profits, together with the additional profits accruing on intermediate products used in that industry. That is so even though the concern is, as in the studies in question, only with misallocation between consumers' goods industries, and not with further losses due to monopolistic distortions in factor mixes. (I return to that later.) If, as a practical expedient, reference is to be made only to "direct" as distinct from "direct and indirect" monopoly profits, it should be closer to the mark to relate such profits to value-added rather than to sales, as Harberger does.

In his calculation Schwartzman seeks to determine the impact of monopoly on *U.S.* industrial prices from a comparison of ratios of sales and variable costs in similar concentrated and unconcentrated industries in the United States and Canada. In this way, the average "monopoly effect" is estimated at 8.3 percent of average variable costs. Schwartzman avowedly seeks to meet objections such as Stigler's to the Harberger calculation. To what extent he succeeds in doing so in his involved calculation is difficult to judge from his very brief exposition, and further questions such as were raised above regarding the Harberger calculation do not seem clearly disposed of. Schwartzman takes a step back from Harberger by assuming a uniform monopoly effect in all monopolized industries. He thus fails to allow for the misallocation due to variations in that effect between products. Monopoly price-cost ratios are a complex matter on which there will have to be many more studies before we can confidently narrow the possibilities.

I referred at the outset to a study of Kamerschen indicating that monopoly losses may be appreciably greater than Harberger and Schwartzman calculated. Kamerschen proceeds for the most part as Harberger did, so his findings too are difficult to construe, but he refers to profits data for a later period, 1956–57 to 1960–61. He also attempts to extend the calculations to embrace nonmanufacturing business and enterprises other than corporations, and to allow more systematically than Harberger could for capitalized monopoly profits and monopolistic distortions in costs in the form of intangible assets, royalties, and advertising expense. While intermediate products are again treated as final, however, monopoly profits are still related to sales rather than value added. Monopoly welfare losses are estimated at 1.03 to 1.87 percent of the national income depending on which of a number of variants is considered. These magnitudes, while modest, are, of course, far larger than Harberger's. They result from the assumption of a unitary elasticity of demand. Alternative calculations apparently assuming elasticities in different industries such as conform to (10) yield still much higher figures: 3.87 to 6.82 percent of national income.[7]

III. An Alternative Approach

In inquiring into the methodology of Harberger, Schwartzman, and Liebenstein, I have elaborated an alternative that is akin to theirs so far as monopoly welfare losses are still calculated from uncompensated demand schedules. But while the alternative methodology served as a basis for appraising that of Harberger, Schwartzman, and Leibenstein, it evidently must be difficult to apply when monopolized products are at all numerous. I now explore still another approach to the calculation of welfare losses due to monopoly pricing. Resting on the assumption of a special type of household indifference map, this approach can only yield hypothetical measures, but these may illuminate further what might already have been surmised: the marked sensitivity of the calculated welfare losses to the elasticity of demand for monopolized products and to the level and distribution of the monopoly price-cost ratios.

The household indifference map in question is given by equation (11):

$$(11)\quad U(x_1, \ldots, x_{n+1}) = \sum_{i=1}^{n+1} A_i x_i^{(1-1/\sigma)}$$

As before, I assume that there are $n+1$ industries producing a corresponding number of commodities, that of these industries one, $n+1$, is competitive and that all others are monopolistic. The output and

[7] I refer to calculations based on after-tax profits. Further calculations based on before-tax profits often yield even higher figures.

hence also the household's consumption of the ith good is x_i.

In (11), A_i is a constant, and so too is σ. If x_i stood for employment of a productive input rather than consumption, therefore, (11) would simply represent on the $x_1 \ldots x_{n+1}$ hyperplane a *CES* production function such as lately has become familiar in productivity analysis. The term σ would then represent the elasticity of substitution between any two factors, employment of all others being given.

I refer here nevertheless to consumption, but as Paul Samuelson, p. 787, has pointed out, the *CES* production function itself corresponds to a function of the same form that I derived long ago (1936, reprinted 1966) for a household's indifference map in an analysis of Ragnar Frisch's methods of measuring marginal utility. In any event, (11) may be conceived of as applying to consumption no less than to production, and for present purposes it has the distinct merit that by varying σ, now representing the elasticity of substitution between any two consumer's goods, we may allow different degrees of substitutability between products, and so ultimately for varying elasticities of demand for one or another of our monopolized goods. In the absence of empirical data on the comparative income elasticities of monopolized and competitive goods, it may be more of a virtue than a limitation of (11) that it implies unitary income elasticities of demand for all products alike.

While I shall refer to $U(x_1, \ldots, x_{n+1})$ as a utility function, our concern is only with the indifference map that the function defines. Where appropriate, utility might be envisaged as being represented by some function of $U(x_1, \ldots, x_{n+1})$, rather than by that function itself. The utility dimension, therefore, need not be representable by a linear homogeneous function such as is usually considered in productivity analysis. Even on that understanding, (11) is admittedly rather restrictive, but it still embraces an interesting range of possibilities. I use a function of the form of (11) here at the suggestion of Samuelson, who has also been helpful at other points.[8]

As before, the monopolized products, $x_1 \ldots x_n$, sell at prices $p_1 \ldots p_n$, which when fixed monopolistically have the values $p_1^m \ldots p_n^m$. The competitive good, x_{n+1}, serves as numeraire. I again designate by I^c the household's income as it would be if $x_1 \ldots x_n$ were to sell at competitive prices, $p_1^c \ldots p_n^c$, and there were no unemployment. Then

$$GCV = I^* - I^c \tag{12}$$

where

$$U(p_1^m, \ldots, p_n^m, I^*) = U(p_1^c, \ldots, p_n^c, I^c) \tag{13}$$

In other words, I^* is the income that the household would require if, when $x_1 \ldots x_n$ sell at $p_1^m \ldots p_n^m$, it is to enjoy the same utility as it would when $x_1 \ldots x_n$ sell at $p_1^c \ldots p_n^c$ and its income is I^c.

We also have

$$MIA = I^m - I^c \tag{14}$$

where I^m is the income that the household

[8] As he has also pointed out to me, the usual *CES* variant of (11), i.e., the one that is linear homogeneous, might itself be properly taken as a measure of real income. Moreover, for shifts along a linear production possibility schedule such as I have been assuming and will continue to assume here, real income as so understood and the net compensating variation turn out to be essentially the same metric. Thus, I express the *NCV* below as a coefficient, the decisive term being the ratio of two magnitudes of household income, one representing the income needed to assure, under monopoly pricing in monopolistic industries, the same utility as might have been enjoyed under competitive pricing in all industries, and the other representing simply the income needed to assure full employment under monopoly pricing in monopolistic industries. That ratio, as is not difficult to see, corresponds precisely to the ratio of real income in equilibrium under competitive pricing in all industries to real income in equilibrium under monopoly pricing in monopolistic industries.

must have if employment is to continue to be full when prices are $p_1^m \ldots p_n^m$ rather than $p_1^c \ldots p_n^c$. Finally,

$$(15) \quad NCV = (I^* - I^c) - (I^m - I^c) = I^* - I^m$$

While NCV is of interest, we may more readily relate our analysis to previous ones if I focus on a corresponding coefficient:

$$(16) \quad CNCV = (I^* - I^m)/I^m$$

In effect, this expresses the NCV as a fraction of the national income that would be produced in the full employment equilibrium under monopoly pricing.

If $x_i, i=1 \ldots n$, is measured in units such that $p_i^c=1$, production possibilities, which are again linear, are readily expressed in terms of a familiar formula convenient here:

$$(17) \quad \sum_{i=1}^{n+1} x_i = k$$

From (11), (16), (17) and the usual conditions for household equilibrium, it can be proven that:

$$(18) \quad CNCV = \left(\sum_{i=1}^{n+1} \gamma_i \lambda_i^{\sigma-1} \right)^{1/(\sigma-1)} \cdot \left(\sum_{i=1}^{n+1} \gamma_i \lambda_i^{-1} \right) - 1$$

Here, for $i=1 \ldots n$, λ_i is p_i^m when the unit of x_i is such that $p_i^c=1$, and also represents the monopolistic price-cost ratio. The term λ_{n+1} equals unity. Also,

$$(19) \quad \gamma_i = (p_i^m x_i^m)/I^m$$

or since $p_i^m=\lambda_i$

$$(20) \quad \gamma_i = \lambda_i x_i^m/I^m$$

where x^m represents the output of x_i when p_i, $i=1 \ldots n$ is fixed monopolistically, and there is full employment equilibrium, i.e., income is I^m. Thus γ_i is the income share that is accounted for in such an equilibrium by the household's expenditure on x_i. A proof of (18) is given in the Appendix.

From (18), the $CNCV$ depends on: (i) γ_i, the income share devoted to x_i, $i=1 \ldots n+1$, in the full employment equilibrium with monopoly prices; (ii) λ_i, $i=1 \ldots n$, the monopoly price-cost ratio for each and every product, it being understood that λ_{n+1}, the price-cost ratio for the competitive product, is unity; and (iii) σ the elasticity of substitution. In order to see how the $CNCV$ might vary in dependence on these parameters, let us consider first the case where there is but one monopolistic product, x_1, and the competitive product is accordingly x_2. Our parameters, then, consist simply of γ_1, λ_1, and σ. In Table 1, I show in percentages the magnitudes of the $CNCV$ indicated by alternative hypothetical values of γ_1, λ_1, and σ. In all cases, γ_1 is taken equal to 0.5, so the household divides its income equally between the monopolized and competitive products. For λ_1, the monopoly price-cost ratio, I consider three alternative values: 1.1,

Table 1—Values of the Coefficient of Net Compensating Variation ($CNCV$) for a Two-Good Economy and Alternative λ_1 and σ

σ	$CNCV$[a] $\lambda_1=1.1$	$\lambda_1=1.2$	$\lambda_1=1.3$
(1)	(2)	(3)	(4)
.50	.06	.21	.43
1.00	.11	.41	.86
2.00	.23	.83	1.73
4.00	.45	1.66	3.43
8.00	.90	3.20	6.38
16.00	1.70	5.47	9.95
32.00	2.35	7.58	12.46
64.00	3.36	8.30	13.74

[a] Shown in percent.

TABLE 2—VALUES OF THE COËFFICIENT OF NET COMPENSATING VARIATION (*CNCV*) FOR A THREE-GOOD ECONOMY FOR ALTERNATIVE $M_m(\lambda)$, λ_1, λ_2, AND σ

	CNCV[a]					
	$M_m(\lambda)=1.1$		$M_m(\lambda)=1.2$		$M_m(\lambda)=1.3$	
σ	$\lambda_1=1.05$ $\lambda_2=1.15$	$\lambda_1=1.05$ $\lambda_2=1.20$	$\lambda_1=1.1$ $\lambda_2=1.3$	$\lambda_1=1.1$ $\lambda_2=1.4$	$\lambda_1=1.15$ $\lambda_2=1.45$	$\lambda_1=1.15$ $\lambda_2=1.60$
	(1)	(2)	(3)	(4)	(5)	(6)
.50	.08	.10	.28	.33	.54	.62
1.00	.17	.20	.56	.70	1.11	1.29
2.00	.33	.42	1.15	1.43	2.33	2.81
4.00	.67	.88	2.43	3.18	4.98	6.48
8.00	1.41	1.97	5.15	7.39	10.43	15.09
16.00	2.93	4.46	9.82	15.10	17.93	27.08
32.00	5.36	8.46	14.34	21.85	23.38	35.01
64.00	7.50	11.54	16.94	25.47	26.21	39.03

[a] Shown in percent.

1.2, and 1.3. The elasticity of substitution, σ, varies between .50 and 64.[9]

The *CNCV* evidently varies widely in dependence on both these parameters. Thus, for $\lambda_1=1.1$, the *CNCV* varies from less than a tenth of a percent to 3.9 percent as σ varies from .50 to 64. For $\lambda_1=1.2$, the corresponding variation in *CNCV* is from two-tenths of a percent to 8.8 percent, and for $\lambda_1=1.3$, from four-tenths of a percent to 13.7 percent. For the single household economy in question, as indicated, reference in each case is in effect to the percentage relation of monopoly welfare losses to the national income. The *CNCV* is also affected by γ_1, though often not very markedly. Interestingly, depending on σ, a deviation of γ_1 in either direction from 0.5 may raise or lower the *CNCV*.[10]

Turning to many monopolized products, from the calculations just discussed, the *CNCV* evidently must depend on the average λ_i in monopolized industries. By merely increasing the number of monopolized products to two, we see that the *CNCV* is also sensitive to the distribution of λ_i about their mean.

Thus, in Table 2, I assume that there are two monopolized industries, x_1 and x_2, and one competitive one, x_3. The corresponding price-cost ratios are λ_1, λ_2, and λ_3, where $\lambda_3=1$. For the weighted mean of λ_1 and λ_2, we have

$$(21)\quad M_m(\lambda) = \left(\sum_{i=1}^{n} \gamma_i\lambda_i\right) \Big/ \left(\sum_{i=1}^{n} \gamma_i\right)$$

In the table, I consider, for $n=2$, three alternative values of $M_m(\lambda)$ corresponding to the three values of λ_1 that were considered previously, that is, 1.1, 1.2, 1.3. For each $M_m(\lambda)$ I refer to two variants, one in which λ_1 and λ_2 are distributed symmetrically about their mean, and the other in which their distribution about their mean is asymmetric, with λ_2 ex-

[9] Where $\sigma=1$, formula (11) gives way to the familiar Cobb-Douglas variant. In that case it is easy to show by reasoning such as is used to prove (18), that that formula is supplanted by:

$$CNVC = \left(\prod_{i=1}^{n+1} \lambda_i^{a_i}\right)\left(\sum_{i=1}^{n+1} a_i\lambda_i^{-1}\right) - 1$$

[10] For example, for $\lambda_1=1.2$, and σ taken in turn to be 2.0, 8.0, 16.0, and 46.0, I find that for $\gamma_1=.33$ the *CNVC* has these values: .74, 3.21, 6.19 and 11.37. For $\gamma_1=.67$, the corresponding figures are .74, 2.56, 4.04, and 5.98.

ceeding $M_m(\lambda)$ by more than λ_1 falls short of it. The specific λ_1 and λ_2 assumed in each variant are as indicated in the table. In all cases, γ_i for the competitive good—now γ_3—is 0.5, so as with the single monopolized good considered in Table 1, the two monopolized goods account for half the household's income. For the symmetric distributions, then, $\gamma_1=\gamma_2=0.25$. For the asymmetric distributions, $M_m(\lambda)$ is as assumed for $\gamma_1=.33$ and $\gamma_2=.17$.

With the inclusion of a second monopolized good in the economy the *CNCV* apparently increases markedly. That is so for all the assumed $M_m(\lambda)$ and for all the symmetric distributions of λ_1 and λ_2. It is even more so if λ_1 and λ_2 are distributed asymmetrically. For example, for $\sigma=8$, a monopoly price-cost ratio of 1.2 implied previously a *CNCV* of 3.2 percent. With a mean price-cost ratio for the two monopolized industries of 1.2, the *CNCV* now rises to 5.15 percent for the symmetric distribution. For the asymmetric distribution, the coefficient is still greater: 7.39 percent.

Further calculations assuming that the number of monopolistic products is indefinitely large but with price-cost ratios distributed in a more or less skew way, as might be expected, yield for a wide range of σ measures of the *CNVC* very similar to those obtained from the asymmetric distribution in Table 2.[11]

In selecting values of parameters for evaluation of the *CNCV*, I have tried to take account of related magnitudes considered in previous analyses, though I also refer to quite other possibilities as well, though perhaps not always very realistic ones. One of the parameters considered, however, is σ, the elasticity of substitution. That is, of course, not the same thing as the elasticity of demand considered in previous analyses, but for the utility function being considered the two aspects are related by a very simple formula:[12]

$$\eta^i_{xp} = \sigma - \gamma_i(\sigma - 1) \tag{22}$$

Here, η^i_{xp}, as before, is the negative of $(p_i/x_i)(\partial x_i/\partial p_i)$, the partial derivative being understood to have the usual sort of subscripts. Evidently, if $\sigma<1$, then $\eta^i_{xp}>\sigma$, but otherwise η^i_{xp} is no greater than σ, and is in fact less than σ whenever $\sigma>1$. The shortfall in the latter case is the greater the larger is σ and the larger is γ_i, the share of the commodity in question in the household's budget. While theoretically the divergence between η^i_{xp} and σ could thus be marked, presumably γ_i would tend to be small so far as there are many commodities. For practical purposes, therefore, η^i_{xp} should usually closely approximate σ, and my calculations may be read accordingly.

I have assumed that production possibilities are linear. That is more or less the counterpart of the assumption of constant costs made in previous analyses. Where costs are not constant but increasing, it has been held (see Harberger (1954, p. 82)) that other things equal, monopoly welfare losses will be less than where costs are constant. The appropriate translation of increasing costs here is a curvilinear production possibilities schedule that is concave from below. With such a schedule, the structural shift in outputs induced by

[11] The calculations assume as before that the competitive product accounts for one-half of income. The remaining half is distributed among monopolistic products on the assumption that the share of income represented by price cost ratios within a small interval $\Delta\lambda$ embracing any particular ratio λ is given by the product $\Delta\lambda$ and a linear function $\gamma(\lambda)=a\lambda+b$ which slopes downward to the right and reaches zero from above at $\lambda=h\equiv-(b/a)$. The constants a and b are evaluated so as to assure that $M_m(\lambda)$, the mean monopoly λ, corresponds to those considered previously: 1.1, 1.2, and 1.3. To cite one or two examples for $\sigma=8.00$ the *CNCV* for $M_m(\lambda)=1.1$ is 1.96; for $M_m(\lambda)=1.2$, 7.39; for $M_m(\lambda)=1.3$, 15.19.

[12] See the Appendix.

monopoly pricing is necessarily dampened, and one might suppose that as a result monopoly welfare losses would indeed be reduced. That could be so, but I believe it need not be. I leave this matter, however, to separate inquiry. Of course, so far as there is monopoly, we cannot at all exclude that in the vicinity of the equilibrium costs will be decreasing or that production possibilities will be convex from below.

I have also assumed that the elasticity of substitution is the same between any and all pairs of goods. As seen from (22) that need not mean that the elasticity of demand is also the same for all goods, but I have tacitly assumed that monopoly price-cost ratios might vary between goods independently of their price elasticities. As indicated, (10) can hardly be expected to apply systematically, but so far as it applies at all the price-cost ratios must in some degree vary inversely with price elasticities. Given that, one surmises that losses would be less than otherwise, but I leave this, too, to separate inquiry.

VII. Conclusions

I have sought to appraise calculations indicating that welfare losses due to monopoly pricing in a country such as the United States are relatively inconsequential, and have argued that those calculations are open to question on both conceptual and empirical grounds. Under an alternative approach, which does not seem subject to conceptual limitations affecting previous calculations, the coefficient of net compensating variation, which is taken to measure monopoly welfare losses, is quite sensitive to empirical aspects which still seem unsettled, particularly, the elasticity of demand for monopolized products, and the varying magnitudes of price-cost ratios for different monopolized products, together with the shares of household income that those products account for. While I have queried the conclusions of Harberger, Schwartzman, and Leibenstein, they have focused attention in a forceful way on an important question. That is a significant contribution in itself, from which nothing that has been said can detract.

I have joined previous writers in tacitly assuming that in the community considered, supplies of productive factors are given. No account has thus been taken of the adverse effect of monopolistic pricing on choices between leisure and work. The analysis can be extended in familiar ways to embrace such choices, however, so they seem to pose no new question of principle, though since leisure must be treated as a competitive good, its inclusion must affect the share of such goods in household income.[13]

I have also joined previous writers in focusing on welfare losses due to monopolistic pricing as that affects resource allocation as between consumer's goods. Monopolistic pricing, of course, also causes misallocation of resources at other points. Such further misallocation is properly the subject of a separate inquiry, but the calculations in question are sometimes taken to reflect monopoly welfare losses generally. Perhaps I should underscore, therefore, that whatever the losses due to misallocation between consumer goods, they are only compounded by misallocation caused by monopoly prices in other spheres, perhaps chiefly the determination of comparative factor proportions in different industries,[14] and the volume of saving and investment. Related to, but not the same thing as, misallocation due to monopoly pricing is the adverse impact of

[13] Monopolistic pricing also may have an adverse effect on effort, but at least in principle it is not difficult to extend the analysis to that aspect too.

[14] Harberger (1959) has tried to grapple with this matter, too, though reference is to distortions due to not only monopoly prices but other causes.

monopoly on the discovery and introduction of new technologies, though Schumpeter, of course, long ago made it a controversial matter whether monopoly is really always disadvantageous in this sphere. But the point being made is obvious and need not be labored: If consequential welfare losses due to monopoly are not precluded in respect of consumption structure, they hardly are so more generally.

Appendix

This appendix derives equations (18) and (22). To begin with (18), given (11) and that x_{n+1} is numeraire, we derive at once these equilibrium conditions for household consumption:

$$\text{(A1)} \quad U_i/U_{n+1} = (A_i/A_{n+1})(x_{n+1}/x_i)^{1/\sigma} = p_i, \qquad i = 1, \ldots, n$$

Here and elsewhere U_i, $i=1 \ldots n+1$, represents the marginal utility of x_i. Formula (A1) holds whether prices for monopolistic products, i.e., $x_i \ldots x_n$, are at competitive or monopolistic levels. But, given our choice of units for those products, $p_i^c = 1$, $i = 1 \ldots n$. Hence, when prices are at competitive levels and I is such as to assure full employment at those prices, the consumption of x_i and x_{n+1} conform to the relation:

$$\text{(A2)} \qquad \overset{c}{x}_i = (A_i/A_{n+1})^{\sigma} \overset{c}{x}_{n+1}$$

From this and (17),

$$\text{(A3)} \qquad \sum_{i=1}^{n+1} (A_i/A_{n+1})^{\sigma} \overset{c}{x}_{n+1} = k$$

It follows that

$$\text{(A4)} \qquad \overset{c}{x}_{n+1} = kA_{n+1}^{\sigma} \Big/ \sum_{i=1}^{n+1} A_i^{\sigma}$$

and

$$\text{(A5)} \quad \overset{c}{x}_i = kA_i^{\sigma} \Big/ \sum_{i=1}^{n+1} A_i^{\sigma}, \qquad i = 1, \ldots, n$$

When prices of monopolistic products are at monopoly levels, $\lambda_1 \ldots \lambda_n$, and money income is I^m, which assures full employment at these prices, we see by similar reasoning that the consumption of x_{n+1} is:

$$\text{(A6)} \qquad \overset{m}{x}_{n+1} = kA_{n+1}^{\sigma} \Big/ \sum_{i=1}^{n+1} \lambda_i^{-\sigma} A_i^{\sigma}$$

and of x_i,

$$\text{(A7)} \qquad \overset{m}{x}_i = k\lambda_i^{-\sigma} A_i^{\sigma} \Big/ \sum_{i=1}^{n+1} \lambda_i^{-\sigma} A_i^{\sigma}, \qquad i = 1, \ldots, n$$

Consider now the still different equilibrium that prevails when prices of monopolistic products are at monopoly levels, but money income is at I^*, as given by (13). Proceeding again much as before, we find that consumption of x_i and x_{n+1} conform to:

$$\text{(A8)} \qquad \overset{*}{x}_i = \lambda_i^{-\sigma} (A_i/A_{n+1})^{\sigma} \overset{*}{x}_{n+1}$$

In order to prove (18), I derive from the foregoing relations certain formulas for I^* and I^m. To refer first to I^*, we know that

$$\text{(A9)} \qquad I^* = \sum_{i=1}^{n+1} \lambda_i \overset{*}{x}_i$$

Also, as seen from (13),

$$\text{(A10)} \quad U(\overset{*}{x}_1, \ldots, \overset{*}{x}_{n+1}) = U(\overset{c}{x}_1, \ldots, \overset{c}{x}_{n+1})$$

From this and (11),

$$\text{(A11)} \quad \sum_{i=1}^{n+1} A_i (\overset{*}{x}_i)^{(1-1/\sigma)} = \sum_{i=1}^{n+1} A_i (\overset{c}{x}_i)^{(1-1/\sigma)}$$

Using this, (A2), and (A8),

$$\text{(A12)} \quad \overset{*}{x}_{n+1} = \overset{c}{x}_{n+1} \left\{ \sum_{i=1}^{n+1} A_i^{\sigma} \right\}^{\sigma/(\sigma-1)} \Big/ \left\{ \sum_{i=1}^{n+1} \lambda_i^{1-\sigma} A_i^{\sigma} \right\}^{\sigma/(\sigma-1)}$$

and

$$\text{(A13)} \quad \overset{*}{x}_i = \lambda_i^{-\sigma} (A_i/A_{n+1})^{\sigma} \overset{c}{x}_{n+1} \cdot \left\{ \sum_{i=1}^{n+1} A_i^{\sigma} \right\}^{\sigma/(\sigma-1)}$$

$$\Big/ \left\{ \sum_{i=1}^{n+1} \lambda_i^{1-\sigma} A_i^{\sigma} \right\}^{\sigma/(\sigma-1)}$$

From (A4), (A9), and (A13),

$$\text{(A14)} \quad I^* = k \left\{ \sum_{i=1}^{n+1} A_i^{\sigma} \right\}^{1/(\sigma-1)} \Big/ \left\{ \sum_{i=1}^{n+1} \lambda_i^{1-\sigma} A_i^{\sigma} \right\}^{1/(\sigma-1)}$$

Turning to I^m, we have

$$\text{(A15)} \quad I^m = \sum_{i=1}^{n+1} \lambda_i x_i^m$$

Using (A7),

$$\text{(A16)} \quad I^m = k \left\{ \sum_{i=1}^{n+1} \lambda_i^{1-\sigma} A_i^{\sigma} \right\} \Big/ \left\{ \sum_{i=1}^{n+1} \lambda_i^{-\sigma} A_i^{\sigma} \right\}$$

From (16), (A14), and (A16),

$$\text{(A17)} \quad CNCV = \left[\left\{ \sum_{i=1}^{n+1} A_i^{\sigma} \right\}^{1/(\sigma-1)} \left\{ \sum_{i=1}^{n+1} \lambda_i^{-\sigma} A_i^{\sigma} \right\} \Big/ \left\{ \sum_{i=1}^{n+1} \lambda_i^{1-\sigma} A_i^{\sigma} \right\}^{\sigma/(\sigma-1)} \right] - 1$$

From (A7) and (A16),

$$\text{(A18)} \quad \gamma_i \equiv (\lambda_i x_i^m / I^m) = \left(\lambda_i^{1-\sigma} A_i^{\sigma} \Big/ \sum_{i=1}^{n+1} \lambda_i^{1-\sigma} A_i^{\sigma} \right)$$

Formula (18) follows from (A17) and (A18).

While it may have been useful to prove (18) in the foregoing manner, it should be observed that the same result can be obtained somewhat more quickly if it is considered that from (16) the *CNVC* reduces to $(I^*/I^m - 1)$. Also, as given by (11) the utility function is homogeneous. Hence, equilibrium consumption positions relating to different levels of money income but the same prices all must be on the same ray through the origin. That is true particularly of the equilibrium consumption positions corresponding to I^* and I^m. It also follows that the ratio I^*/I^m corresponds to the comparative consumption of any one good in the two equilibrium positions. Given that, (18) follows from (A13), (A7), (A4), and (A16).

To come to (22), from (A1),

$$\text{(A19)} \quad p_j^{\sigma}(\partial x_j/\partial p_i) - (A_j/A_{n+1})^{\sigma} \cdot (\partial x_{n+1}/\partial p_i) = 0, \qquad j \neq i,$$

and

$$\text{(A20)} \quad p_i^{\sigma}(\partial x_i/\partial p_i) - (A_i/A_{n+1})^{\sigma} \cdot (\partial x_{n+1}/\partial p_i) = -\sigma p_i^{\sigma-1} x_i$$

In these formulas, the derivatives shown are of a partial sort; with appropriate subscripts understood. Also, since

$$\text{(A21)} \quad \sum_{i=1}^{n+1} p_i x_i = I,$$

we have

$$\text{(A22)} \quad \sum_{j=1}^{n+1} p_j(\partial x_j/\partial p_i) = -x_i$$

From (A19), (A20), and (A22),

$$\text{(A23)} \quad (\partial x_{n+1}/\partial p_i) = (\sigma - 1) A_{n+1}^{\sigma} x_i \Big/ \sum_{i=1}^{n+1} p_i^{1-\sigma} A_i^{\sigma}$$

And from this and (A20),

$$\text{(A24)} \quad (\partial x_i/\partial p_i) = \left\{ (\sigma - 1) A_i^{\sigma} x_i / p_i^{\sigma} \sum_{i=1}^{n+1} p_i^{1-\sigma} A_i^{\sigma} \right\} - \sigma x_i / p_i$$

or

$$\text{(A25)} \quad \eta_{xp}^{i} = \sigma - (\sigma - 1) p_i^{1-\sigma} A_i^{\sigma} \Big/ \sum_{i=1}^{n+1} p_i^{1-\sigma} A_i^{\sigma}$$

Formula (22) follows at once from (A18) for monopoly equilibrium where $\lambda_i = p_i$, but it is evident that it must hold for any other equilibrium as well, for in (A18) λ_i could be construed as well to represent any p_i, and not merely a monopolistic one.

REFERENCES

A. **Bergson**, "Real Income, Expenditure Proportionality and Frisch's *New Methods*," *Rev. Econ. Stud.*, Oct. 1936, *9*, 33–52.

———, *Essays in Normative Economics*, Cambridge, Mass. 1966.

A. C. **Harberger**, "Monopoly and Resource Allocation," *Amer. Econ. Rev. Proc.*, May 1954, *44*, 77–87.

———, "Using the Resources at Hand More Effectively," *Amer. Econ. Rev. Proc.*, May 1959, *49*, 134–46.

———, "Three Basic Postulates for Applied Welfare Economics: An Interpretive Essay," *J. Econ. Lit.*, Sept. 1971, *9*, 785–97.

J. R. **Hicks**, *A Revision of Demand Theory*, Oxford 1956.

H. **Hotelling**, "The General Welfare in Relation to Problems of Taxation and of Railway and Utility Rates," *Econometrica*, July 1938, *6*, 242–69.

D. R. **Kamerschen**, "Welfare Losses from Monopoly," *Western Econ. J.*, summer 1966, *4*, 221–36.

H. **Leibenstein**, "Allocative Efficiency vs. 'X-Efficiency'," *Amer. Econ. Rev.*, June 1966, *56*, 392–415.

P. A. **Samuelson**, "Using Full Duality to Show That Simultaneous Additive Direct and Indirect Utilities Implies Unitary Price Elasticity of Demand," *Econometrica*, Oct. 1965, *33*, 781–96.

D. **Schwartzman**, "The Burden of Monopoly," *J. Polit. Econ.*, Dec. 1960, *68*, 627–30.

G. J. **Stigler**, "The Statistics of Monopoly and Merger," *J. Polit. Econ.*, Feb. 1956, *64*, 33–40.

Devaluation, Money, and Nontraded Goods

RUDIGER DORNBUSCH*

This paper develops a monetary approach to the theory of currency devaluation.[1] The approach is "monetary" in several respects. The role of the real balance effect is emphasized and a distinction is drawn between the relative prices of goods, the exchange rate and the price of money in terms of goods. Furthermore, money is treated as a capital asset so that the expenditure effects induced by a monetary change are spread out over time and depend on the preferred rate of adjustment of real balances.[2] The latter aspect gives rise to the analytical distinction between impact and long-run effects of a devaluation.

The first part of this paper develops a one-commodity and two-country model of devaluation. The simplicity of that structure is chosen quite deliberately to emphasize the monetary aspect of the problem as opposed to the derivative effects that arise from induced changes in relative commodity prices. Trade is viewed as the exchange of goods for money or a means of redistributing the world supply of assets. A devaluation is shown to give rise to a change in the level of trade and the terms of trade, the price of money in terms of goods.

In the second part the implications of the existence of nontraded goods are investigated, and induced changes in the relative prices of home goods enter the analysis.

I. Devaluation in a One-Commodity World

In this part we develop a purely monetary approach to devaluation in discussing a two-country, two-monies, and one-commodity model.[3] This stripped down model abstracts from the complexities of distribution and substitution effects that may arise from changes in relative commodity prices and places primary emphasis on the real balance effect.

A. *The Model*

We assume that money is the only marketable asset and that real income (output) is in fixed supply in each country. The demand for nominal balances in each country is assumed to have the Cambridge form.[4]

* University of Rochester. This paper draws on my dissertation and I am indebted to the members of my thesis committee, Harry Johnson, Stanley Fischer, and Robert Mundell. In revising the material I had the benefit of comments from Karl Brunner, George Borts, Stanley Engerman, and Murray Kemp. I am particularly indebted to Ronald W. Jones and Michael Mussa with whom I enjoyed extended discussion of the topic.

[1] This approach is by no means novel. For formal developments see Frank Hahn, Jones (1971), Kemp (1969, 1970), Mundell, and Takashi Negishi (1972). Acceptance of that approach has nevertheless remained limited.

[2] A "capital-theoretic" approach to the real balance effect is developed by Alvin Marty.

[3] The notion of trade in one commodity may alternatively be interpreted as trade in a composite commodity, so that relative goods prices remain unchanged. Such conditions may obtain either because of perfect substitution or else because of the absence of distribution effects.

[4] The particular functional form of the demand for money obviously lacks generality. It is chosen here in order not to detract from the main line of argument. Alternative specifications would assume the demand for money proportional to expenditure as in Jones (1971) or else derive the demand for money from intertemporal utility maximization. Provided the underlying utility function is separable in consumption and real balances the qualitative conclusions of this paper carry over to such a formulation.

$$(1) \qquad L = kP\bar{y}; \qquad L^* = k^*P^*\bar{y}^*$$

where

k, k^* = the desired ratios of money to income
$\bar{y}, \bar{y}^*$ = real outputs
P, P^* = the money price of goods in terms of domestic and foreign currency

and where an asterisk denotes the foreign country. Given the exchange rate, e, the domestic currency price of foreign exchange, arbitrage ensures that

$$(2) \qquad P = P^*e$$

With respect to monetary policy we assume that the nominal quantity of money in each country M, M^*, is initially given and that governments abstain from changing domestic money supplies except as it is necessary to maintain a pegged exchange rate. Accordingly the rate of increase in the domestic money supply is given by the trade balance surplus, B.

$$(3) \qquad \dot{M} = B = -e\dot{M}^*$$

Desired nominal expenditure in each country, Z, Z^*, is equal to money income less the *flow* demand for money, H, H^*, where the latter is assumed proportional to the *stock* excess demand

$$(4) \qquad Z = P\bar{y} - H$$
$$Z^* = P^*\bar{y}^* - H^*$$

$$(5) \qquad H = \pi(L - M) = H(P, M);$$
$$H^* = \pi^*(L^* - M^*) = H^*(P^*, M^*)$$

and where π and π^* are the domestic and foreign rates of adjustment. The expenditure functions in (4) imply a short-run marginal propensity to spend out of income smaller than unity while in the long-run, when monetary stock equilibrium is attained, the average propensity to spend equals unity.

In Figure 1 we show the domestic rate of hoarding, H, and the foreign rate of *dis*hoarding, $-H^*$, as a function of P the domestic currency price of goods. The schedules are drawn for given nominal money supplies in each country and an exchange rate e^0. With the nominal quantity of money given, hoarding in the home country is an increasing function of the price level. An increase in the price level creates a stock excess demand for money and causes expenditure to decline relative to income as the community attempts to restore the real value of cash balances. It follows that we may view the hoarding schedule alternatively as the flow demand for money or the excess supply of goods (in nominal terms). By the same reasoning the foreign rate of *dis*hoarding, given the exchange rate, is a decreasing function of the home price level. We note that the distribution of the money supplies underlying Figure 1 is not compatible with balance of payments equilibrium. Foreign monetary stock equilibrium would obtain at P' while for domestic monetary equilibrium the price level would have to be equal to P''.

Consider now the conditions of short-run equilibrium. In order for the world goods market to clear, we require that world income equal world expenditure or

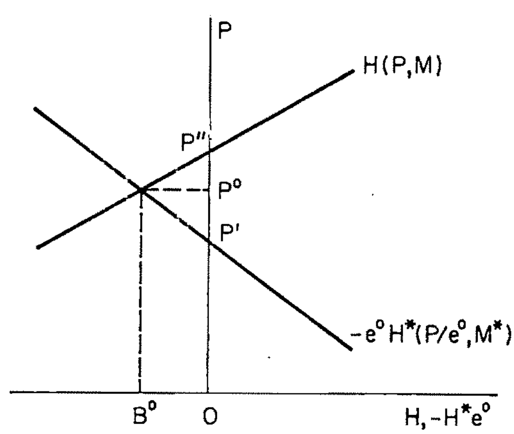


FIGURE 1

equivalently that the home country's rate of hoarding equal the foreign country's rate of dishoarding.

$$H = -H^*e^0 \tag{6}$$

The equilibrium is shown in Figure 1 at a domestic currency price of goods P^0; a higher price level would leave a world excess supply of goods and a lower price level a world excess demand for goods. We observe, too, that the short-run equilibrium at P^0 implies a trade balance deficit for the home country equal to B^0. That deficit, in the absence of sterilization, as we assume, redistributes money from the home country to the rest of the world. The reduction in the domestic nominal quantity of money reduces real balances at the initial price level and thereby causes planned hoarding to decrease and conversely abroad. In terms of Figure 1 this implies that the hoarding and dishoarding schedules shift to the right, a process that continues over time until they intersect between P'' and P' on the vertical axis. At that time exchange of money for goods ceases since each country has achieved its preferred asset position and spends at a level equal to its income.

B. *The Short-Run Effects of a Devaluation*

Consider now the short-run or impact effect of a devaluation on the part of the home country. A devaluation changes the equilibrium relationship between price levels in the two countries. Differentiating equation (2) we obtain

$$\hat{P} = \hat{P}^* + \hat{e} \tag{7}$$

where a ˆ denotes a relative change in a variable. Equation (7) informs us only about the relationship between changes in the price levels at home and abroad; we have to investigate the equilibrium condition in the world goods market in order to determine what the actual change in the price level in each country will be. For that purpose we turn to Figure 2 where we show the world economy in initial long-run equilibrium at a domestic currency price of goods P^0.

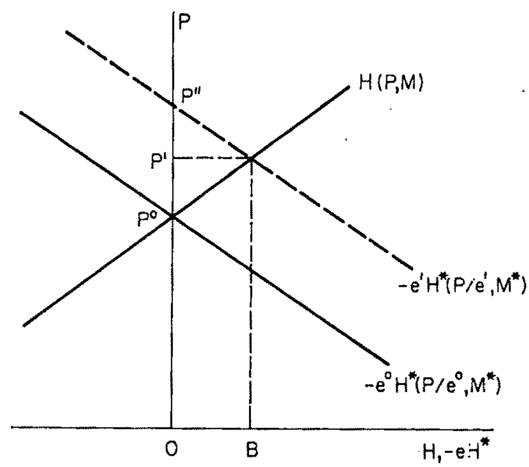


FIGURE 2

The effect of a devaluation is shown in Figure 2 by an upward shift in the foreign dishoarding schedule. For foreign monetary stock equilibrium to obtain, given the nominal quantity of money, the foreign currency price of goods would have to remain constant which in turn by (7) implies that the domestic price level would have to increase in the same proportion as the exchange rate, a price change equal to $(P''-P^0)/P^0$. The domestic hoarding schedule, on the contrary, is unaffected and domestic monetary stock equilibrium would continue to obtain at a domestic price level P^0. It is observed from Figure 2 that at an unchanged domestic price level there would be a world excess demand for goods due to the increase in foreign real balances and expenditure while at an unchanged foreign price level there would be a world excess supply of goods due to the decrease in domestic real balances and expenditure. It follows that in order for the world goods market to clear the price level changes will have to

be distributed in such a manner as to reduce domestic absorption and increase foreign absorption by an equal amount.

The equilibrium increase in the domestic price level is equal to $(P'-P^0)/P^0$ while the foreign price level declines in the proportion $(P''-P^0)/P^0$. We note that both the domestic and foreign currency price of goods change less than proportionately to the rate of devaluation and that the distribution of price changes depends on the relative slopes of the hoarding schedules.

Given these price changes foreign real balances have increased and the real value of domestic balances has decreased thereby causing foreigners to dishoard in order to decumulate their capital gains and domestic residents to save in order to restore the real value of their cash balances. The home country's balance of payments surplus is equal to OB and causes a redistribution of the world money supply.

The formal criterion for the price changes and the balance of payments can be developed by differentiating the goods market equilibrium condition

$$(6') \quad \pi(kP\bar{y} - M) + e\pi^*(k^*P\bar{y}^*/e - M^*) = 0$$

with respect to P and e holding the nominal quantity of money constant in each country. The relative change in the domestic price level is

$$(8) \qquad \hat{P} = \frac{\pi^*M^*e}{\pi M + \pi^*M^*e}\,\hat{e}$$

Defining the world money supply, measured in terms of domestic currency $\overline{M}$,

$$(9) \qquad \overline{M} = M + eM^*$$

and the domestic and foreign country's share in the money world supply, σ and σ^*, we can rewrite (8) as

$$(8') \qquad \hat{P} = \frac{\pi^*\sigma^*}{\pi\sigma + \pi^*\sigma^*}\,\hat{e} \geq 0$$

Substituting (8′) in (7) we obtain the effect of a devaluation on the foreign price level:

$$(10) \qquad \hat{P}^* = \frac{-\pi\sigma}{\pi\sigma + \pi^*\sigma^*}\,\hat{e} \leq 0$$

Equations (3′) and (10) show the distribution of price changes to depend on relative effective size where effective size is the product of the speed of adjustment and the share in the world money supply. In the small country case ($\pi\sigma/\pi^*\sigma^*=0$) the home country price level increases in the same proportion as the exchange rate.

The home country's trade balance surplus is obtained by differentiating the flow demand function for money with respect to the price level and substituting (8) to yield

$$(11) \quad dB = dH = \pi M\left[\frac{\pi^*M^*e}{\pi M + \pi^*M^*e}\right]\hat{e} > 0$$

Equation (11) confirms that the balance of payments unambiguously improves.

C. *The Long-Run Effects of Devaluation*

The long-run effects of devaluation on nominal money supplies and price levels may be interpreted with the help of Figure 3. In quadrants II and IV we show the domestic and foreign demand for real balances as hyperbolae; quadrant III shows the equilibrium price relationship $P^*e=P$ for the initial exchange rate as the ray OA. Lastly in quadrant I the world money supply at the initial exchange rate is given by $\overline{M}\overline{M}^*$, where $\overline{M}^*=\overline{M}/e$.

Initial long-run equilibrium is indicated by point E where the distribution of the world money supply is such that each country holds the desired quantity of real balances and where the equilibrium relationship between price levels is satisfied.[5]

A devaluation on the part of the home

[5] For a similar geometric treatment, see Arnold Collery.

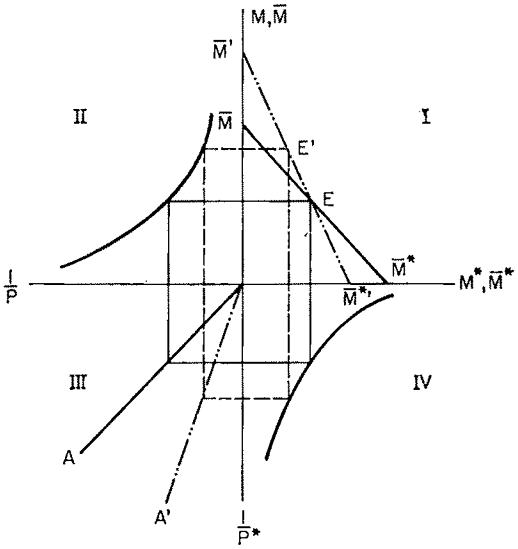


FIGURE 3

country affects both the price relationship and the world money supply. To each domestic price level corresponds now a lower equilibrium foreign price level; this is indicated in quadrant III by a rotation of the arbitrage line to OA'. Furthermore, given the initial nominal quantities of money in each country indicated by point E, the world money supply measured in terms of either currency changes: it decreases when measured in terms of foreign currency by the initial domestic quantity of money times the exchange rate change and it appreciates in terms of domestic currency by the initial foreign quantity of money times the exchange rate change. Accordingly the world monetary constraint rotates around point E—the initial endowment of currencies—to become $\overline{M}'\overline{M}^{*\prime}$.

It is readily verified from Figure 3 that the initial distribution of money supplies at point E is no longer appropriate as a long-run equilibrium position since it would be inconsistent with the new price relationship. The new long-run equilibrium is shown by point E' indicating an increased domestic quantity of money and price level and a decreased foreign quantity of money and price level. Real balances obviously remain unchanged, between the new equilibrium and the old.

We should emphasize that our assumption about the absence of national money supply changes other than by the balance-of-payments mechanism is only one possible assumption about the behavior of money supplies. If we had assumed on the contrary that the home country accompanied the devaluation by an equiproportionate increase in its nominal quantity of money the only short- and long-run effect of the combination of policies would be an equiproportionate increase in the domestic price level and no effect whatsoever abroad.[6]

The latter monetary assumption would be appropriate if the home country wished to run a transitory budget deficit financed by money creation without impairing its foreign exchange position; the former assumption corresponds to the case where a country uses a devaluation to increase its foreign exchange holdings.

II. Devaluation and Nontraded Goods

In this part we consider an extension of the monetary model to introduce flexibility in relative prices. Following Jones (1972), Michael Michaely, Mundell, and Anne Krueger, we assume that there are two classes of goods produced and consumed in each country, traded goods and nontraded goods. Each class of goods itself is taken to be a composite commodity so that the relative prices of goods within each group are invariant. The aggregation chosen here places emphasis on the relative price of nontraded in terms of traded

[6] In terms of Figure 3 the policy combination would imply that the world monetary constraint both rotates and shifts outward passing through $\overline{M}^*$ since the world money supply measured in terms of foreign currency would remain unchanged. The conclusions in the text are readily verified from the fact that the new equilibrium point would lie vertically above point E.

goods rather than on the terms of trade between internationally traded goods; it emphasizes the effects of changes in absorption on relative prices rather than the income effect of changes in the relative prices of traded goods.

This extension has two implications for the effects of a devaluation: changes in hoarding or equivalently changes in expenditure relative to income change the equilibrium relative price of home goods and these changes in relative prices in turn affect the equilibrium rates of hoarding.

We will show that in this more disaggregated structure the conclusions of the one-commodity model continue to hold for the effects of a devaluation on the balance of payments and the prices of traded goods; the additional element that arises is that the reduction in domestic absorption and the increase in foreign absorption cause the relative price of home goods to decline at home and to rise abroad. This result may be viewed as a special case of transfer analysis and arises in that perspective since each country's marginal propensity to spend on foreign home goods is by definition zero.[7]

A. *The Model*

Denoting traded and nontraded commodities as goods one and two, respectively, we assume that production takes place along a concave transformation curve and that supplies are a function only of the relative price:

$$X_i = X_i(q) \qquad i = 1, 2 \tag{12}$$

where q is the relative price of nontraded goods—the ratio of the domestic currency prices of nontraded and traded goods, P_2 and P_1, respectively:

$$q = P_2/P_1 \tag{13}$$

Demand for the two commodities is assumed to depend on money prices and nominal expenditure, or, using the homogeneity property and adopting traded goods as a numeraire, on relative prices and real expenditure measured in terms of traded goods, $\tilde{Z}$.[8]

$$C_i = C_i(q, \tilde{Z}) \qquad i = 1, 2 \tag{14}$$

Real expenditure is defined as real income less real hoarding, all measured in terms of traded goods as a numeraire:

$$\tilde{Z} = \tilde{Y} - \tilde{H} \tag{15}$$

where real income or the real value of output is defined as follows:

$$\tilde{Y} \equiv X_1 + qX_2 = \tilde{Y}(q) \tag{16}$$

Monetary considerations affect the goods markets via the expenditure function and in particular via the planned rate of hoarding. Maintaining our assumption that the demand for nominal balances is proportional to money income and that hoarding is proportional to the stock excess demand for money we may write the desired real rate of hoarding, measured in terms of traded goods, as a function of the relative price and the real quantity of money measured in terms of traded goods:

$$\tilde{H} = \tilde{H}(q, \tilde{M}) \tag{17}$$

where

$$\tilde{M} = M/P_1 \tag{18}$$

Our assumptions about the stock demand for money ensure that an increase in either domestic currency price raises the desired rate of real hoarding so that the following properties hold:

$$q\frac{\partial \tilde{H}}{\partial q} \equiv \alpha > 0; \qquad -\tilde{M}\frac{\partial \tilde{H}}{\partial \tilde{M}} \equiv \beta > 0 \tag{19}$$

[7] The relationship between the transfer problem and devaluation is more extensively analyzed in Dornbusch (1973) and Jones (1971).

[8] In the remainder of this paper a tilde will denote the fact that a quantity is measured in terms of traded goods. When these quantities are referred to as "real" this will not imply measurement in terms of a price index.

The definition of real expenditure in (15) may be rewritten as the budget constraint in a manner that reveals the disaggregation of the model:

$$(20)\quad q(X_2 - C_2) + (X_1 - C_1) = \tilde{H}$$

It is evident from the budget constraint that when the home-goods market clears ($X_2 = C_2$) the excess supply of traded goods identically equals the planned rate of hoarding.

Given a corresponding set of behavioral relations and constraints for the foreign country we can now turn to the conditions of short-run equilibrium in this model. Short-run equilibrium obtains when for a given exchange rate and given money supplies, all goods markets clear; that is, when the market for nontraded goods clears in each country and when the world market for traded goods clears. Such an equilibrium, by the budget constraint in each country, implies that one country's planned rate of hoarding equals the other country's planned rate of dishoarding. Equations (21) formally state these equilibrium conditions of the model

$$(21)\qquad \begin{aligned} E_2 &\equiv X_2(q) - C_2(q, \tilde{Z}) = 0 \\ E_2^* &\equiv X_2^*(q^*) - C_2^*(q^*, \tilde{Z}^*) = 0 \\ &\tilde{H}(q, \tilde{M}) + \tilde{H}^*(q^*, \tilde{M}^*) = 0 \end{aligned}$$

where

$$\tilde{M}^* \equiv M^*/P_1^*; \qquad q^* \equiv P_2^*/P_1^*; \qquad P_1^* e = P_1$$

The first two conditions in (21) state that in equilibrium the excess demand for home goods is zero in each country while the third equation is the market clearing condition in the market for traded goods.

B. *The Impact Effect of a Devaluation*

To examine the modifications in the effects of devaluation brought about by the introduction of nontraded goods we consider first the relationship between the relative price of home goods and real hoarding. In particular we want to show that an increase in real hoarding lowers the relative price of home goods. That result obtains since an increase in real hoarding represents a decrease in real expenditure relative to real income so that at constant relative prices and given a positive marginal propensity to spend on home goods the demand for home goods decreases. A decline in the relative price of home goods is required in order to eliminate the excess supply generated by an increase in hoarding. More formally the relationship between the relative price of home goods and real hoarding may be derived by differentiating the first market equilibrium condition in (21) to obtain

$$(22)\quad \hat{q} = -\frac{m_2}{(\eta_2 + e_2)qC_2}\, d\tilde{H}$$

where

$$m_2 \equiv q\frac{\partial C_2}{\partial \tilde{Z}} > 0$$

$$\eta_2 \equiv -\frac{q}{C_2}\left[\frac{\partial C_2}{\partial q} + \frac{\partial C_2}{\partial \tilde{Z}}\,\frac{\partial \tilde{Y}}{\partial q}\right] > 0$$

$$e_2 \equiv \frac{\partial X_2}{\partial q}\,\frac{q}{X_2} > 0$$

The terms m_2, η_2, and e_2, denote respectively, the marginal propensity to spend on home goods, the compensated elasticity of demand for home goods, and the elasticity of supply.

In Figure 4 we show the market equilibrium schedule for the home country's nontraded goods market as the locus $E_2 = 0$; to maintain market equilibrium the expenditure reducing effects of an increase in hoarding have to be offset by the substitution effects of a decrease in the relative price of home goods.

So far we have treated hoarding as the exogeneous variable and have enquired into the relative price effects of changes in

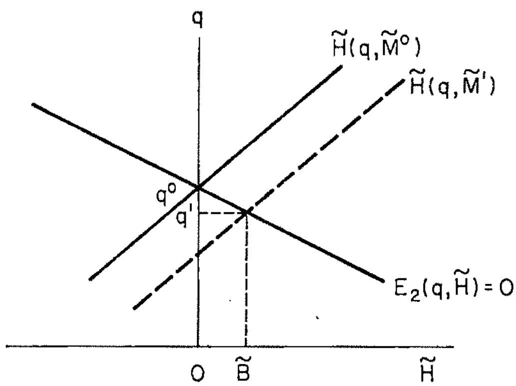


FIGURE 4

hoarding. We wish next to develop an expression that relates the rate of hoarding, given the nominal quantity of money, to price changes. Differentiating the hoarding function in (17) we obtain

$$d\tilde{H} = \alpha\hat{q} + \beta\hat{P}_1 \tag{23}$$

and substituting for the change in the relative price of home goods, $\hat{q}$, from (22), we obtain

$$d\tilde{H} = \gamma\beta\hat{P}_1 \tag{24}$$

where the terms

$$\gamma \equiv \frac{1}{1 + \alpha\delta} > 0; \qquad \delta \equiv \frac{m_2}{(\eta_2 + e_2)qC_2} > 0$$

are introduced for notational convenience. To gain further understanding of the relationship between hoarding, relative prices and the money price of traded goods derived in (24) we turn to Figure 4 where we show the effect of an increase in the price of traded goods. In addition to the market equilibrium schedule for home goods we draw a hoarding schedule as an increasing function of the relative price of home goods, given the nominal quantity of money and the price of traded goods and hence the real quantity of money, $\tilde{M}^0$. The schedule is upward sloping since an increase in the price of home goods raises income and hence the demand for money thereby increasing the desired rate of hoarding.

Initial equilibrium is shown at a relative price of nontraded goods, q^0. An increase in the price of traded goods reduces the real money supply and hence increases at constant relative prices the desired rate of hoarding. This is shown in Figure 4 by a rightward shift of the hoarding schedule. Since at constant relative prices there is an excess supply of home goods, their relative price will decline to q' which in turn dampens the equilibrium rate of hoarding, $\tilde{B}$, relative to what it would have been at constant relative prices. The shift in the hoarding schedule corresponds to the term $\beta\hat{P}_1$ in (24) while the dampening effect shows in the term γ.

It will be recognized that in the composite commodity model analyzed earlier perfect substitutability ensured that $\delta=0$. In the present formulation the absence of perfect substitution and the requirement that home-goods markets clear ensure that absorption changes are reflected in changes in relative prices; furthermore these induced changes in relative prices affect the equilibrium rate of hoarding tending to reduce the hoarding response associated with a given change in the price of traded goods.

Having developed the basic relationships of the model we can now proceed to investigate the effects of a devaluation. For that purpose we turn to Figure 5. In the upper part of that figure we draw the domestic and foreign home goods market equilibrium schedules, where the latter is drawn as a function of the foreign rate of *dis*hoarding and hence is negatively sloped. We assume, arbitrarily and without consequence, that initially the relative prices of home goods are the same in both countries. In the lower part of Figure 3 we draw the domestic hoarding schedule and the foreign dishoarding schedule. It is important to note that along these hoarding schedules the relative price of home goods

is allowed to adjust in order to clear the home-goods market so that by the budget constraint these hoarding schedules may alternatively be interpreted as the domestic excess supply of traded goods and the foreign excess demand for traded goods. Analytically the schedules are defined by equation (24) and its counterpart for the foreign country.

Initial equilibrium obtains at a domestic currency price of traded goods P_1^0 and equilibrium relative prices of home goods $q^0=q_0^*$. A devaluation by the home country may be analyzed in a manner similar to the composite commodity model developed above. At an unchanged domestic currency price of goods foreign real balances increase causing foreigners to dishoard which is shown in Figure 5 by a

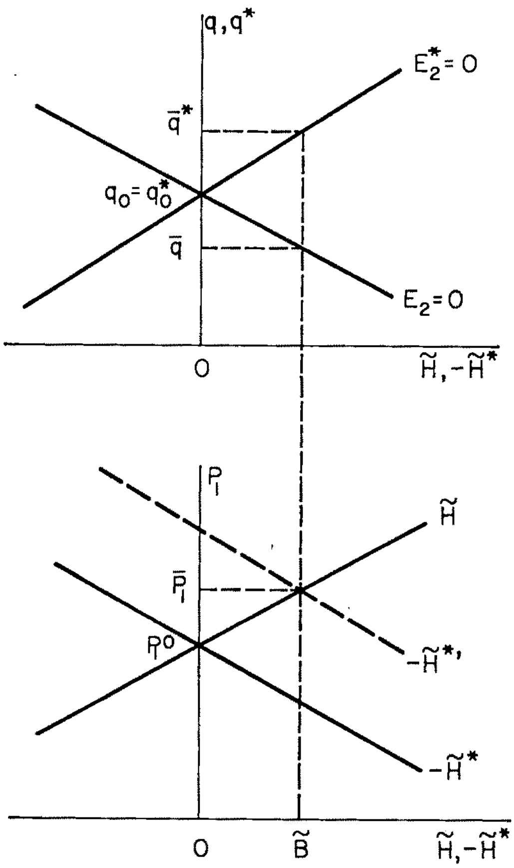


FIGURE 5

rightward shift of the foreign dishoarding schedule to $\tilde{H}^{*\prime}$. Short-run equilibrium will obtain at a domestic currency price of goods $\bar{P}_1$ where the world market for traded goods clears. The increase in the domestic price of traded goods causes the home country to reduce expenditure relative to income and run a trade balance surplus equal to $\tilde{B}$. Corresponding to the reduction in domestic absorption we find a decline in the relative price of nontraded goods at home to $\bar{q}$ while the increase in foreign absorption raises the relative price of nontraded goods in that country to $\bar{q}^*$.

These results can be derived more formally by consideration of the equilibrium in the world market for traded goods. Recalling that (24) allows explicitly for market clearing in the home-goods market that expression is identically equal to the excess supply of traded goods. Accordingly we may use (24) and its counterpart for the foreign country to determine the effects of a devaluation on the domestic currency price of traded goods:

$$(25)\quad \beta\gamma\hat{P}_1 + \beta^*\gamma^*(\hat{P}_1 - \hat{e}) = 0 = d\tilde{H} + d\tilde{H}^*$$

Solving for the relative change in the domestic currency price of traded goods yields

$$(26)\qquad \hat{P}_1 = \frac{\beta^*\gamma^*}{\beta\gamma + \beta^*\gamma^*}\,\hat{e} \equiv \theta\hat{e}$$

The solution for the effect of a devaluation on the domestic currency price of traded goods shows that this price will increase less than proportionately to the rate of devaluation $(0<\theta<1)$. Differentiating the price relationship $P_1^*=P_1/e$ it is seen that the foreign currency price of traded goods will fall less than proportionately to the rate of devaluation. Substitution of (26) in (24) shows that the devaluing country's balance of payments unambiguously improves:

$$(27)\qquad d\tilde{H} = \beta\gamma\theta\hat{e}$$

So far our results correspond qualitatively to those obtained in the composite commodity model. The departure arises from the fact that changes in absorption in the two countries change the equilibrium relative prices of home goods. Substituting (27) in (22) we find that a devaluation lowers the relative price of nontraded goods in the home country and raises it abroad:

$$(28) \quad \hat{q} = -\delta\beta\gamma\theta\hat{e}; \quad \hat{q}^* = \delta^*\beta^*\gamma^*(1-\theta)\hat{e}$$

While, like in the first part of this paper, short-run equilibrium is characterized by an exchange of traded goods for real balances and hence the absorption effects of a devaluation are emphasized, the role of the relative price of home goods is nevertheless crucial in the adjustment mechanism. Given imperfect substitutability between home goods and traded goods on the production side it is the adjustment in the relative price of home goods that translates changes in absorption into an excess supply of traded goods at home and an excess demand for traded goods abroad.

III. Concluding Remarks

Rather than summarize here the conclusions of this paper we wish to emphasize some of the issues raised by the present formulation of devaluation analysis.

The first and primary issue concerns the role of money in models of devaluation. The stance taken here is that a devaluation is foremost a monetary phenomenon and that its effects derive from the reduction in the real value of money attendant upon a devaluation. If it is believed that the effects of a reduction of real balances on expenditure, by whatever transmission mechanism, are negligible then it may stand to reason that the effects of a devaluation are negligible not that there must be other powerful avenues through which it exerts its effects.

The second issue that deserves attention is that of aggregation. The formulation developed here suggests that it is helpful to view traded goods as a composite commodity and thus to highlight the distinction between money and goods and between classes of goods that are respectively traded and nontraded.

REFERENCES

A. Collery, *International Adjustment, Open Economies and the Quantity Theory of Money*, Princeton Studies in International Finance, no. 28, Princeton 1971.

R. Dornbusch, "Aspects of a Monetary Theory of Currency Depreciation," unpublished doctoral dissertation, Univ. Chicago 1971.

———, "Currency Depreciation, Hoarding and Relative Prices," *J. Polit. Econ.*, July/Aug. 1973, *81*, 893–915.

F. Hahn, "The Balance of Payments in a Monetary Economy," *Rev. Econ. Stud.*, Feb. 1959, *26*, 110–25.

R. W. Jones, "A Role for Money in the Exchange Model," unpublished, Univ. Rochester 1971.

———, "Alternative Models of Devaluation," unpublished, Univ. Rochester 1972.

M. Kemp, "The Balance of Payments and the Terms of Trade in Relation to Financial Controls," *Rev. Econ. Stud.*, Jan. 1970, *37*, 25–31.

———, *The Pure Theory of International Trade and Investment*, Englewood Cliffs 1969, ch. 14.

O. A. Krueger, "The Role of Home-Goods in Exchange Rate Adjustments," unpublished 1971.

A. Marty, "The Real Balance Effect: An Exercise in Capital Theory," *Can. J. Econ. Polit. Sci.*, Aug. 1964, *30*, 360–67.

M. Michaely, "Relative-Prices and Income-Absorption Approaches to Devaluation: A Partial Reconciliation," *Amer. Econ. Rev.*, Mar. 1960, *50*, 144–47.

R. A. Mundell, *Monetary Theory*, Pacific Palisades 1971, ch. 9–11.

T. Negishi, *General Equilibrium Theory and International Trade*, Amsterdam 1972, 15–17.

The Effect of the EEC and EFTA on European Trade: A Temporal Cross-Section Analysis

By NORMAN D. AITKEN*

Utilizing a cross-sectional trade flow model of the type developed by Hans Linnemann and Jan Tinbergen, this study attempts to isolate empirically the major forces which have shaped European trade relations over the period 1951–67. We first estimate via the use of dummy variables the impact of the European Economic Community (*EEC*) and the European Free Trade Association (*EFTA*) on member trade. For each year of the European integration period (1959–67), a cross-sectional equation is estimated and used to test for the existence and approximate size of the respective integration effects. The equation is also calculated for the eight years prior to the integration period to obtain a clear picture of the forces which were at work before the formation of the *EEC*. Secondly, a base year equation is used to make projection estimates of the gross trade creation and European trade diversion effects of the two communities.

I. Methodology

As defined by Bela Balassa (1967, p. 5), gross trade creation (*GTC*) will refer to the total increase in trade among members of a trading community brought about through integration, regardless of whether the additional trade replaces domestic production or whether it replaces nonmember exports. The substitution of imports from member countries (higher cost imports) for imports from nonmember countries (lower cost imports) will constitute trade diversion (*TD*). External trade creation (*ETC*) will refer to integration-caused increases in trade between a trading community and countries outside the trading community. *ETC* as an effect of economic integration is possible in the case of a customs union where high tariff countries have to reduce their tariffs to the outside world as a part of the process of achieving a common external tariff. *ETC* minus *TD* yields the net effect of a trading bloc on the outside world.

All empirical studies which have attempted to measure integration effects have been faced with the common problem of isolating the effect of integration on trade from the effect of income growth and changes in other variables which normally affect international trading patterns. The major approaches to this problem have either been to examine changes in market share of imports (or apparent consumption)[1] or to incorporate income directly into the statistical analysis (by

* Associate professor of economics, University of Massachusetts. Research for the paper was supported in part by a grant from the Research Council of the University of Massachusetts. I am indebted to Ronald Ehrenberg, Bradley Gale, James Kindahl, Joseph Kushner, Laurence Mauer, Anthony Scaperlanda, George Treyz, Brendan Walsh, anonymous referees, and the editor for valuable comments on earlier drafts of the paper. I assume all responsibility for any errors that remain.

[1] Various formulations of the market-share approach are used in the studies by Edwin Truman, P. S. Verdoorn, and F. J. M. Meyer Zu Schlochtern, J. Waelbroeck, M. K. Carney, and The European Free Trade Association (1969).

calculating income elasticities of import demand for the pre- and postintegration periods[2] or by using income as an independent variable in a trade-estimating model).[3] Both of these approaches attempt to measure the effects of integration indirectly, or in other words as a residual. Either by using a base year trade matrix or a base period relationship between income and imports, estimates are made of what trade would have been in the absence of economic integration and these estimates are then compared to actual trade to obtain the trade preference effect. One is thus left with the uncertainty as to whether or not other factors may have been responsible for the difference between projected and actual values. Furthermore, the residual approach does not provide an objective basis for determining the first year in which integration effects occurred. The base periods selected by previous studies of the *EEC* range from 1958 to 1961 and hence it is clear that there is no general agreement as to the timing of the first *EEC* effect on trade.

The present study has attempted to deal with these problems by estimating, through the least squares regression method, the following variant of the Linnemann trade flow model for each year of the 17-year period 1951–67.

$$(1)\quad \log X_{ij} = \log b_0 + b_1 \log D_{ij} + b_2 \log Y_i + b_3 \log Y_j + b_4 \log N_i + b_5 \log N_j + b_6 \log A_{ij} + b_7 \log P_{ij}^{EEC} + b_8 \log P_{ij}^{EFTA} + \log e_{ij}$$

where X_{ij} is the dollar value of country i's exports to country j measured according to country j's import data, Y_i and Y_j are the nominal GNP of countries i and j expressed as a dollar value, N_i and N_j are the populations of the respective countries, D_{ij} is the distance between the commercial centers of the two countries, A_{ij} is a dummy variable for adjacent or neighboring countries, P_{ij}^{EEC} and P_{ij}^{EFTA} are dummy variables for trade between partners of the *EEC* and *EFTA*, respectively, and *log* refers to common logarithms.[4]

The variables Y_i and N_i together determine the potential export supply of country i, with Y_i determining economic capacity and N_i determining the domestic market/foreign market production ratio. (N determines market size and, assuming economies of scale, the larger N the more lines of production for which the country will meet the minimum market size for efficient market production (see Linnemann, pp. 11–14). The larger N, therefore, the larger the domestic market to foreign market ratio and the smaller the potential export supply of the country.) The variables Y_j and N_j together determine the potential import demand for country j, using the same arguments that were applied to potential export supply. D_{ij} is a proxy variable for natural trade resistence which in turn is a composite of transportation cost, transport time, and economic horizon. Consequently, D_{ij} along with N_i and N_j is hypothesized to have a negative effect on X_{ij}.

Neighboring countries (A_{ij}) can be expected to have an additional stimulus to trade because of similarity of tastes and an awareness of common interests (see Balassa (1961, p. 40)). Perhaps more important, however, neighboring countries are likely to experience significant additional amounts of international trade in

[2] See Bela Balassa (1967).

[3] See Mordechai Kreinin and J. Waelbroeck.

[4] Sources of data: *Direction of Trade; Distance Between Ports; EFTA Trade: 1969; International Financial Statistics; National Accounts of OECD Counties;* and *Monthly Bulletin of Statistics.* Values of 2 and 1 were used for the three dummy variables to indicate the presence and absence of the given characteristic. A mimeographed data appendix is available from the author.

the form of what are essentially locally traded goods, especially where border regions are densely populated as in much of Europe.[5]

The use of this model permits us to incorporate into the analysis as independent variables the preference area effects through the use of dummy variables. Dummy variables are used to represent in an approximate way phenomena which are difficult to measure. The approximate nature of the trade preference variables means that their measured effect on trade flows must in turn be considered approximate. On the other hand, there are distinct advantages to the use of the cross-sectional model.

First, by estimating the preference effect as an independent variable in a multiple regression equation one is able to hold constant other major variables which affect trade, including not only potential demand and supply but also to some extent the effects of general changes in trade liberalization and transportation cost. Changes in the latter over time would tend to be picked up by the proxy variables D_{ij} and A_{ij}. (The proxy variables themselves, of course, do not change over time, but changes in the real variables would be expected to show up as changes from year to year in the estimated coefficients of the proxy variables.) Specifically, since the size of the trade-stimulating effect of European trade liberalization may be expected to vary inversely with distance (natural trade resistance) the effect of trade liberalization may show up as a change over time in the D_{ij} or A_{ij} coefficients or both.

In addition, the use of a cross-sectional model allows the estimation of trade preference coefficients for each year in the integration period and hence a series of parameter estimates can be obtained which can then be considered as a whole in terms of whether their pattern indicates the expected cumulative growth in the preference effects.

The preference parameters can in turn be used to estimate the dollar value of *GTC* for each of the two European communities. Because the estimates for each year are derived from the cross-sectional equation for that year, each estimate is independent of the others and the estimating procedure does not require the use of a base year. In fact, the results of the study provide information which may be useful in determining when the first integration effects on trade occurred.

Using the trade-preference coefficients to estimate the trade-stimulating effect of integration in any postintegration year requires one to assume that the size of the coefficient is being determined solely by the effect of the trade preference and therefore that it is not in part reflecting some other special trade relationship which had existed in the preintegration period. Consequently, it will be necessary to test the preference coefficients for nonsignificance in the preintegration period as well as for significance in the postintegration period. By calculating the equation for the eight years preceding the first tariff reductions of the *EEC*, we were thus able to test for

[5] Equation (1) represents what Tinbergen has labeled a "turnover equation" in which prices are not specified (see Linnemann pp. 4, 47 for a detailed discussion of this point). Variables which explicitly measure artificial trade barriers are also absent from the equation because of the overwhelming statistical and methodological problems associated with their direct measurement (Linnemann, p. 30). The exclusion of these variables from the equation could be expected to seriously bias the trade preference coefficients only if there were a high correlation between the excluded variables and the trade preference variables themselves. If a high correlation existed, however, it would show up in the form of significant or near significant trade preference coefficients in the preintegration years. Since, as will be seen below, the preintegration preference coefficients are consistently insignificant, it seems reasonable to assume that the preference variables are not highly correlated with any important variable that has been excluded from the statistical analysis.

the existence of preintegration preference effects.

The sample from which the equations were calculated includes the original seven *EFTA* members plus the five *EEC* trading countries (Belgium-Luxembourg being one trading country).[6] The sample thus contains 20 trade flows between *EEC* partners, 42 trade flows between *EFTA* partners, and 70 trade flows between the members of the two trade blocs for a total of 132 observations per year. The 70 trade flows between the two blocs are assumed to constitute "normal" European trade and it is against these normal trade flows that trade among members of the respective communities is tested for preferential effects.[7] Clearly, this assumption would be invalid for any year in which trade flows between the two blocs have, on the average, experienced significant amounts of *TD*.[8] Previous studies, however, have found little or no net trade diversion by either bloc. But rather than rely on the findings of previous studies, it was decided to test for the existence of *TD* between the two blocs by the following procedure: The cross-section regression results were examined to find an appropriate base year free of integration effects. The equation for the base year was then used to estimate what trade would have been between the two blocs in subsequent years had there been no integration in Europe. A comparison of the projection estimates with actual trade provides an estimate of the degree to which trade between the two blocs has been reduced as a result of trade diversion. The projection approach was also used to estimate gross trade creation within the *EEC* and *EFTA* as a further check on the dummy variable estimates. Consequently, the second part of the statistical analysis consists of residual estimates of the dollar value of gross trade creation and trade diversion, but they are residual estimates based on the information provided by the regression results as to the timing of the first *EEC* effect on European trade. Furthermore, the assumption of normal trade between the *EEC* and *EFTA* should be considered only as an initial working hypothesis to be tested against the projection estimates before any conclusions are reached concerning the general magnitude of *GTC* within the two trading communities.

II. Empirical Results

A. *Regression Results*

Table 1 contains the estimated parameter values for the trade flow equation for the 17-year period 1951–67. The trade

[6] See Table 3 for a listing of the individual countries. Finland was excluded from the sample because its late entry into *EFTA* meant that its inclusion in the sample would have made it more difficult to discern the first *EFTA* effect on the original members.

[7] The decision to limit the sample only to members of the two trading communities may be open to question. Ideally, one would prefer to have "normal trade" represented by the exports of both *EFTA* and *EEC* members to countries who are not members of either bloc. Within Western Europe, however, only less developed countries fall into the nonmember category and the possibility of having normal trade represented by trade flows to the developed countries outside of Europe was rejected because a significant proportion of the observations on normal trade would have occupied extremely high points on the regression lines with respect to the independent variables, while all intra-*EEC* and intra-*EFTA* observations would have been in the lower range. It was concluded, therefore, that a sample limited to the specified European countries would provide the greatest opportunity for obtaining bias-free preference coefficients in the preintegration years and hence that one would obtain the most sensitive measure for identifying both the timing and general magnitude of the preference effects. The fact that trade flows between the two blocs had to be used to represent normal trade precluded the introduction of additional dummy variables to measure the potential trade diversion effect (*TD*) of each bloc against the other.

[8] The existence of *TD* would lead to an inflated estimate of *GTC* since the value of the trade preference coefficients would tend to be increased by an integration-caused reduction in the average trade flow between the two blocs as well as by an integration-caused increase in the average trade flow among members. It should be noted, however, that while the existence of *TD* could lead to an inflated estimate of *GTC*, it could not result in the finding of a *GTC* effect where none exists since by definition *TD* cannot occur unless *GTC* occurs as well.

TABLE 1—REGRESSION EQUATIONS FOR EUROPEAN TRADE FLOWS

Year	Constant	Coefficients of Independent Variables[a]								$\bar{R}^2$	S.E.
		D_{ij}	Y_i	Y_j	N_i	N_j	A_{ij}	P_{ij}^{EEC}	P_{ij}^{EFTA}		
1951	1.958	−.427	1.137	1.000	−.493	−.476	.480	−.141	.053	.766	.277
		2.53	*8.81*	*7.76*	*3.60*	*3.48*	*1.85*	*.46*	*.23*		
1952	2.130	−.499	1.163	.876	−.567	−.350	.490	−.213	−.078	.760	.292
		2.84	*8.92*	*6.72*	*3.99*	*2.46*	*1.80*	*.66*	*.32*		
1953	2.155	−.509	1.200	.839	−.599	−.350	.515	−.130	−.068	.769	.292
		2.89	*9.28*	*6.49*	*4.17*	*2.44*	*1.89*	*.40*	*.28*		
1954	2.052	−.484	1.110	.816	−.482	−.354	.686	−.146	−.111	.755	.306
		2.63	*8.27*	*6.08*	*3.23*	*2.37*	*2.40*	*.43*	*.44*		
1955	1.915	−.452	1.081	.810	−.463	−.312	.759	−.091	−.098	.777	.291
		2.58	*8.35*	*6.26*	*3.20*	*2.16*	*2.80*	*.28*	*.41*		
1956	2.000	−.476	1.075	.773	−.482	−.276	.742	−.098	−.107	.782	.284
		2.80	*8.54*	*6.14*	*3.41*	*1.96*	*2.80*	*.31*	*.46*		
1957	1.900	−.448	1.118	.776	−.505	−.304	.788	−.019	−.136	.779	.292
		2.57	*8.64*	*5.99*	*3.48*	*2.09*	*2.89*	*.06*	*.56*		
1958	1.901	−.444	1.069	.740	−.481	−.281	.766	−.008	−.159	.770	.291
		2.55	*8.18*	*5.66*	*3.25*	*1.90*	*2.82*	*.02*	*.66*		
1959	1.848	−.449	1.123	.849	−.484	−.396	.758	.204	−.109	.802	.276
		2.71	*8.91*	*6.74*	*3.44*	*2.81*	*2.94*	*.68*	*.47*		
1960	1.617	−.383	1.215	.903	−.578	−.429	.782	.402	−.103	.815	.273
		2.34	*9.65*	*7.17*	*4.09*	*3.04*	*3.07*	*1.33*	*.45*		
1961	1.617	−.398	1.209	.826	−.551	−.375	.798	.475	.045	.831	.257
		2.59	*10.30*	*7.03*	*4.17*	*2.84*	*3.33*	*1.68*	*.21*		
1962	1.562	−.410	1.150	.925	−.474	−.456	.806	.527	.076	.846	.251
		2.75	*10.18*	*8.18*	*3.72*	*3.58*	*3.45*	*1.91*	*.36*		
1963	1.589	−.440	1.106	.891	−.441	−.376	.789	.580	.172	.854	.244
		3.06	*10.03*	*8.08*	*3.55*	*3.02*	*3.47*	*2.17*	*.85*		
1964	1.520	−.444	1.120	.959	−.439	−.442	.747	.630	.326	.870	.229
		3.28	*10.88*	*9.31*	*3.80*	*3.83*	*3.51*	*2.52*	*1.73*		
1965	1.349	−.392	1.108	.899	−.421	−.396	.828	.743	.345	.863	.232
		2.87	*10.62*	*8.62*	*3.64*	*3.42*	*3.84*	*2.94*	*1.81*		
1966	1.252	−.389	1.111	.880	−.398	−.354	.825	.802	.435	.871	.225
		2.93	*10.88*	*8.62*	*3.54*	*3.15*	*3.94*	*3.26*	*2.35*		
1967	1.067	−.349	1.052	.911	−.331	−.369	.892	.887	.572	.874	.217
		2.74	*10.39*	*9.00*	*3.03*	*3.38*	*4.41*	*3.75*	*3.21*		

[a] X_{ij} is the dependent variable; all variables are expressed in *logs*; *t*-values shown in italics, where 1.66 and 2.36 are significant at the .05 and .01 level, respectively.

preference coefficients come very close to fitting the expected theoretical pattern. In all the preintegration years (1951–58) the P^{EEC} coefficient is not significantly different from zero and even has a negative sign.[9] In 1959, the first year of integration, there is a sharp increase in the value of the coefficient (at least relative to earlier years) as the coefficient becomes positive

[9] Despite the consistent negative sign in the preintegration years there is a small perceptible movement in the P^{EEC} coefficient in 1957 and 1958. This small movement could be reflecting the effect of any of a number of different underlying factors, including the French devaluation, the European Coal and Steel Community, or a possible lagged effect of European trade liberalization. It seems unlikely, however, that any of these factors could have been responsible for the large changes in the coefficient which occur during the integration period. Since rapid progress was made in the liberalization of intra-European trade during the early 1950s (see Robert Triffin, ch. 5, and the *OEEC*, pp. 31–32), one would expect the resulting trade stimulation to be reflected in the regression equations during the early years of the study period. The A_{ij} coefficient exhibits a definite increasing trend from 1951 through 1957 after which it tends to stabilize at around 0.8 until near the end of the study period where it shows another small increase in value. The behavior of the A_{ij} coefficient, therefore, is more consistent with the expected timing and magnitude of intra-European trade liberalization than is the smaller and later preintegration movement of the P^{EEC} coefficient.

for the first time. The value of the coefficient continues to increase in subsequent years, reaching a significance level of .1 in 1960 and becoming statistically significant at the .05 level in 1961. The change in sign and large change in value of the P^{EEC} coefficient (relative to earlier years) from 1958 to 1959 is consistent with the hypothesis that the first *EEC* effect on member trade occurred in 1959, but the hypothesis of no *EEC* effect cannot be rejected at the standard .05 confidence level until 1961.[10] In dealing with the methodological issue of selecting a base year for making projection estimates, however, the appropriate methodological question is not whether the null hypothesis of no *EEC* effect can be rejected at the .05 level, but rather whether or not the hypothesis of no *EEC* effect can be accepted. To make the latter decision, the hypothesis to be tested for acceptance (i.e., no *EEC* effect) becomes the alternative hypothesis and the hypothesis that the *EEC* has had a positive effect on trade becomes the equivalent of the null hypothesis (i.e., the hypothesis to be tested for rejection). Clearly, the positive P^{EEC} coefficients for 1959 and 1960 and the change in sign and large increase in value from 1958 to 1959 do not permit the rejection of the hypothesis of a positive *EEC* effect and hence the hypothesis of no *EEC* effect cannot be accepted for those years. The year 1958, therefore, constitutes the last date for which the regression results allow one to assume that there was no *EEC* effect on member trade.[11]

The coefficient for the *EFTA* preference area follows a pattern similar to that of the *EEC*. In the preintegration years (1951 through 1959) the coefficient is insignificant and in all but the first year it carries a negative sign. In 1960, the first year of integration, the sign is still negative, but since the tariff reductions took place in July of that year, there would have been very little time for the new tariff levels to affect trade flows. The coefficient becomes positive (although it is extremely small) in the following year, grows slowly through 1963, and becomes statistically significant (.05 level) in 1964. The behavior of the P^{EFTA} coefficient is thus consistent with the hypothesis that the first *EFTA* effect on trade took place in 1961, although the null hypothesis of no *EFTA* effect cannot be rejected at the .05 confidence level until 1964.[12] All the other estimated parameters in the equations have the correct sign and all are significant at at least the .05 level in every year.

[10] The standard t-test for the significance of the coefficient constitutes only part of the statistical evidence which should be considered in evaluating the null hypothesis for 1959 and 1960, since it does not take into consideration all the information provided by the total set of yearly regression results. A runs test on the yearly sign of the P^{EEC} coefficient for the 17-year period yields a probability of .00008 that the sign pattern could have occurred by chance. A test of the sign of the yearly change in the value of the coefficient, to test whether or not the coefficient is changing in a random manner, gives $P>.97$ that the coefficient is changing in a random manner for the preintegration period (1951–58), but $P<.001$ (ten consecutive increases) for the integration period (1958–67).

[11] Studies which have used a base year beyond 1958 assume that the *EEC* could not have affected trade in 1959 and/or 1960 either because the 1959 tariff cuts were extended to all *GATT* members or because of an assumed lag in the response of firms to the tariff reductions. The 1959 internal tariff cuts of the *EEC*, however, were specified in the Treaty of Rome, signed in March 1957, while the decision to extend the 1959 tariff reductions to non-*EEC* members was made by the Council of Ministers only on Dec. 3, 1958. Furthermore, the extension of the 1959 tariff reductions to *GATT* members was subject to the important proviso that no tariff would be reduced below the eventual common external tariff (*CET*) for the product. Since the tariffs of Germany and the Benelux countries were already below the *CET*, the tariff reductions to other *GATT* members were limited, for the most part, to French and Italian imports. (See the *Three Banks Review*.) It is by no means clear, therefore, that there could not have been an *EEC* effect on member trade as early as 1959 and certainly the question is not above empirical investigation.

[12] $P=.0007$ that the sign pattern of the P^{EFTA} coefficient could have occurred by chance. $P>.92$ that the coefficient is changing in a random manner for the preintegration period (1951–59); $P<.01$ that change is random during the integration period.

B. *Dummy Variable and Projection Estimates*

The trade preference coefficients provide a measure of the factor by which normal trade among members has been increased as a result of the formation of *EEC* and *EFTA*. Estimates of *GTC* for each trading community, therefore, can be derived from the coefficients for each year of the respective integration periods.[13] These estimates are presented and discussed below in conjunction with the projection estimates.

As noted above, 1958 is the last year for which it can safely be assumed that there was no *EEC* effect on European trade. Accordingly, 1958 was selected as the base year for making the projection estimate of what trade would have been in the absence of economic integration.

In order to make the projections, the 1958 equation was recalculated leaving out the two trade preference variables and the following results were obtained:

$$(2)\quad \log X_{ij} = 1.978 - \underset{(3.76)}{.487} \log D_{ij} + \underset{(8.33)}{1.062} \log Y_i + \underset{(5.75)}{.733} \log Y_j - \underset{(3.36)}{.459} \log N_i - \underset{(1.90)}{.259} \log N_j + \underset{(2.84)}{.718} \log A_{ij} + \log e_{ij}$$

$$\bar{R}^2 = .776 \qquad S.E. = .289$$

The t-values are all significant above the .05 level.[14]

Equation (2) cannot account for the trade effects of changes in competitive position among the countries in the sample[15] nor can it account for any general trade liberalization effects which may have occurred after 1958. The projection estimates are made, therefore, on the basis of the usual assumption that the effect of these factors on trade has been small relative to the effects of integration. For the estimation of *GTC*, this assumption is supported by the regression results of the present study, where the pronounced increase in the trade preference coefficients indicates that economic integration provided the major impetus for additional trade (with the effect of income held constant) among the respective members of the two blocs. For trade between the two blocs, however, the model cannot provide information as to the general magnitude of the *TD* (or *ETC*) effect.

The exponential form of equation (2) was used to estimate intra- and inter-*EEC* and *EFTA* trade and the estimated trade values were then subtracted from actual trade to obtain the residual estimates of the trade effects of the *EEC* and *EFTA*. Since equation (2) estimates the dollar value of trade in 1958 prices, estimated trade was multiplied by country i's export

[13] Since raw data values of 1 and 2 were used for the dummy preference variables, $2^{\hat{b}_7}$ provides an estimate of the factor by which the average trade flow among *EEC* members has been increased as a result of integration. Actual intra-*EEC* trade divided by $2^{\hat{b}_7}$, therefore, yields an estimate of what intra-*EEC* trade would have been in the absence of integration. This value in turn, is subtracted from actual trade to obtain the estimated *GTC* of the *EEC* for the given year.

[14] The error term was included in the equation because the cross-section residuals tended to be stable over time. Real *GNP* in 1958 prices was used to make the projection estimates since it provided the best measure of how the economic capacity of each country has changed relative to the base year. A detailed explanation of these points is provided in an extended mimeographed copy of the paper.

[15] The fact that there was significant dispersion in the export price trends of the individual countries within each trading community over the projection period (see *International Financial Statistics*, Aug. 1968, p. 30) means that the projection estimates for each country are likely to be subject to error, but it also means that there will be at least some tendency for the errors to cancel out in the aggregate results for each community as a whole (i.e. not all competitive effect errors will be in the same direction). Consequently, less confidence can be placed in the individual country estimates than in the aggregate community results. Even in the latter case, however, the results should be considered only as indications of approximate magnitude.

TABLE 2—NET EFFECTS OF INTEGRATION ON *EEC* AND *EFTA* TRADE—
DUMMY VARIABLE AND PROJECTION ESTIMATES
(Millions of Dollars at Current Prices)

	Net *EEC* Effect on:			Net *EFTA* Effect on:		
	EEC Exports[a]		*EFTA* Exports[b]	*EFTA* Exports[a]		*EEC* Exports[c]
Year	Dummy Variable Estimate[d]	Projection Estimate[e]	Projection Estimate[e]	Dummy Variable Estimate[d]	Projection Estimate[e]	Projection Estimate[e]
1959	1,067	925	50	0[f]	−8	66
1960	2,468	1,639	31	0[f]	140	48
1961	3,284	2,254	67	126	149	−102
1962	4,114	3,213	393	222	243	−201
1963	5,203	4,731	541	545	389	−262
1964	6,388	5,695	202	1,151	573	−289
1965	8,228	6,941	− 41	1,326	690	−259
1966	9,784	8,612	−157	1,773	919	−205
1967	11,127	9,189	−629	2,425	1,264	−202

[a] Estimates of Gross Trade Creation (*GTC*).
[b] Estimates of the net external trade creation (*ETC*) or trade diversion (*TD*) effect of the *EEC* on the exports of *EFTA* countries.
[c] Estimates of the net *TD* effect of *EFTA* on the exports of *EEC* countries.
[d] Dummy variable estimates of *GTC* derived from the P^{EEC} and P^{EFTA} coefficients reported in Table 1.
[e] Actual trade in current prices minus trade estimated by equation (2) converted to current prices.
[f] Zero values are given for 1959 and 1960 since the P^{EFTA} coefficient is negative for those years.

price index (dollar prices, 1958 base) in order to obtain estimates of the effects of integration in current prices.[16] The projection estimates therefore can be compared directly with the dummy variable (*DV*) estimates which reflect the prices of the given year for which the regression equation was calculated. The projection estimates for each of the four subgroupings of the sample are contained in Table 2 along with the *DV* estimates of *GTC*.

Because the economic integration of both trading communities has been a cumulative process (i.e., internal tariffs being reduced by stages over the integration periods covered by the study), one should expect to find the estimates of *GTC* increasing from year to year with no reversals. Both the *DV* and projection estimates of the respective communities' effect on member trade are consistent with this expectation. For the *EEC*, the projection estimates are consistently below the *DV* estimates, with the gap increasing at the end of the study period. Despite this discrepancy, both estimates show a large *EEC* effect occurring in 1959 and a strong cumulative growth in the yearly values of *GTC*. For *EFTA*, both estimates also show a cumulative growth in *GTC*, but the *DV* estimates are below (but reasonably close to) the projection estimates through 1962. Starting in 1963, however, the *DV* estimate is greater, with the gap increasing progressively in subsequent years. Even if allowance is made for a large margin of error, the results clearly indicate that the *GTC* of the *EEC* has been substantially greater than that of

[16] The use of total export price indexes introduces error into the results since composition of exports to the *EEC* and *EFTA* may differ from total export composition (although each country's exports to the respective trading communities would, in general, cover a broad spectrum of its total commodity exports).

EFTA. In percentage terms, the projection estimate of *GTC* accounts for 38 percent of actual intra-*EEC* trade (import data) and 16 percent of total *EEC* exports in 1967. For *EFTA*, the projection estimate of *GTC* for 1967 accounts for 16 percent of intra-*EFTA* trade and only 4 percent of total *EFTA* exports.

In considering the effects of each trading community on the other, *EFTA* could only be expected to have a negative effect on *EEC* exports since there was no reduction in the external tariffs of *EFTA* countries. Furthermore, the size of the *TD* effect should increase progressively throughout the integration period since increasing trade discrimination against nonmembers would result from the progressive reduction in internal tariffs. The estimated *EFTA* effect on *EEC* exports, therefore, is only partially consistent with theoretical expectation. The fact that the estimated net effect becomes negative in 1961 (the first full year of integration) and the existence of increasing negative values through 1964 is consistent with the *TD* hypothesis, while the indicated reduction in the size of the *TD* effect during the last three years of the period is not. Disaggregation of the *EFTA* effect on *EEC* exports by individual *EFTA* importing country revealed that the decrease in the "*TD* effect" over the specified period was accounted for almost entirely by the United Kingdom, with the net effect on *EEC* exports increasing from −\$38 million in 1963 to \$530 million in 1967. Since the period 1964–67 coincides with the general decline in the *U.K.* balance of trade position culminating in devaluation at the end of 1967, the divergence of the *U.K.* effect can be attributed to the serious deterioration of the *U.K.* competitive position. The net effect on *EEC* exports of all *EFTA* countries other than the United Kingdom over the same period was found to be consistent with the expected increasing *TD* effect (the estimated net *TD* effect increased progressively from −\$223 million in 1963 to −\$731 in 1967).

Because of the reductions in external tariffs by the original high tariff *EEC* countries (i.e., France and Italy), both *ETC* and *TD* are theoretically possible effects of the *EEC* on *EFTA* trade. The *TD* effect would be expected to dominate eventually, however, since the reduction of internal tariffs toward zero would eventually result in trade discrimination against nonmembers, even in the markets of the original high tariff countries. The estimates of the net *EEC* effect on *EFTA*, therefore, are not inconsistent with theoretical expectation, since they show a general increasing net *ETC* effect through 1963,[17] followed by the emergence of a growing net *TD* effect over the last three years of the period.

Finally, the fact that *TD* is dominating trade between the two communities over the latter years of the period explains the growing divergence between the *DV* and projection estimates for these years. It can be concluded, therefore, that the *DV* estimates are definitely inflated for the years 1965–67 and possibly for earlier years as well. Hence, the *DV*-estimates must be rejected in favor of the projection estimates for at least the last three years of the period.

C. *Individual Country Estimates*

Having found that the results for the two trading communities as a whole are consistent with theoretical expectation, we now proceed to the question of whether the expected results hold for each of the individual countries in the sample for the

[17] Disaggregation of the *EEC* effect by original high and low tariff *EEC* countries provided additional support for the *ETC* hypothesis since the results showed that the increasing *ETC* effect for the period 1959–63 was entirely accounted for by the original high tariff *EEC* countries.

TABLE 3—NET EFFECT OF *EEC* AND *EFTA* ON 1967 TRADE OF INDIVIDUAL COUNTRIES—PROJECTION ESTIMATES

(Millions of Dollars at Current Prices)

Exporting Country	Net *EEC* Effect	Net *EFTA* Effect
Belgium-Lux.	1649	− 39
France	2170	379
Germany	2473	−390
Italy	1958	260
Netherlands	939	−412
Total *EEC*	9189	−202
Austria	−123	234
Denmark	−350	144
Norway	− 30	211
Portugal	− 42	107
Sweden	−100	397
Switzerland	−229	173
United Kingdom	245	− 2
Total *EFTA*	−629	1264

last year of the study period.[18] This represents a more stringent test of the *GTC* and *TD* hypotheses, since there is less opportunity for errors caused by competitive effects to cancel out in the estimates for the individual countries than there is in the aggregated community results. Table 3 contains the individual country projection estimates for 1967. Because of the possibility of error, inter-country comparisons in the size of the estimated effects should be avoided. The results should be viewed primarily in terms of whether or not they are consistent with the *GTC*, *TD*, or *ETC* hypotheses.[19] With the exception of the small negative value for the United Kingdom, the estimated effects of the *EEC* and *EFTA* on the trade of the respective member countries are all positive and hence consistent with the expected *GTC* effect. (Again, the discrepancy for the United Kingdom can be attributed to the deterioration of the *U.K.* competitive position.)[20]

The estimated net *EEC* effect on the individual *EFTA* countries indicates that all *EFTA* countries other than the United Kingdom were experiencing net *TD* as of 1967.

Disaggregation of the 1967 *EFTA* country estimates by original high and low tariff *EEC* countries showed that the high tariff countries accounted for $238 million of the total *EEC* effect on the United Kingdom and hence the positive effect for the United Kingdom can be attributed to *ETC*.[21] Sweden ($82 million) and Switzerland ($7 million) also showed evidence of an *ETC* effect by the high tariff *EEC* countries, but it was more than offset by a much larger *TD* effect from the low tariff countries.[22] Consequently, the results indicate that while *TD* was the dominant effect of the *EEC* towards *EFTA* in 1967, the three major industrial countries of *EFTA* continued to benefit from *ETC* in the French-Italian market.

The estimated *EFTA* effect on *EEC* countries shows positive values for both France and Italy which are contrary to theoretical expectation, since only a *TD* effect could be expected.[23] While it is pos-

[18] Copies of individual country estimates for earlier years are available from the author.

[19] Because of the large magnitude of the difference in *GTC* estimates between the respective members of the two communities, it is possible to conclude that the *GTC* of each *EEC* member has been greater than the *GTC* of any *EFTA* member.

[20] The estimated *EEC* effect on each of the five member countries was positive for all years of the period 1959–67 and the estimated *EFTA* effect on each of the seven member countries was consistently positive after 1962 with the exception of the United Kingdom in 1967.

[21] While *ETC* by France and Italy explains virtually all of the positive *U.K.* value for 1967, the fact remains that there is no evidence of a net *TD* effect by the low tariff *EEC* countries against the United Kingdom as of 1967. (This was true for earlier years as well.) This finding is surprising in view of the fact that *TD* and the worsening *U.K.* competitive position should both be operating to produce a negative effect.

[22] A net *TD* effect by both high and low tariff *EEC* countries prevailed in the case of the four remaining—less industrialized—countries of *EFTA*.

[23] The net effect of all *EFTA* countries other than the United Kingdom was as follows: Belgium-Luxembourg (−131), France (230), Germany (−539), Italy (125), Netherlands (−416). Consequently, while the removal

sible that these two countries experienced only minimal *TD* by *EFTA*, one certainly cannot rule out the possibility that the estimates may have been largely influenced by competitive effects. Consequently, while the net effect of *EFTA* on the *EEC* as a whole is consistent with the *TD* hypothesis, there are sufficient discrepancies among the individual country estimates to temper this conclusion with reservation.

III. Summary and Conclusion

The empirical findings of the study were found to be generally consistent with the expectations of customs union theory. The results showed that both the *EEC* and *EFTA* have experienced a cumulative growth in *GTC* over their respective integration periods, with the *GTC* of the *EEC* being substantially greater than the *GTC* of *EFTA*. The projection estimates for 1967 placed the size of the *GTC* effect for the *EEC* and *EFTA* at approximately \$9.2 billion and \$1.3 billion, respectively. The estimated effect of *EFTA* on the *EEC* as a whole showed a consistent *TD* effect for the period 1961–67, but there were two positive deviations from the expected negative effects of *EFTA* on the exports of the five individual *EEC* countries. The *EEC* was found to have had a net *ETC* effect on *EFTA* through 1964, but this was replaced by a growing net *TD* effect from 1965 through 1967. The estimated effects of the *EEC* on both member-country trade and *EFTA*-country exports were found to be consistent with theoretical expectation.

The regression results of the study yielded the important methodological conclusion that 1958 is the last year for which it can safely be assumed that European trade was unaffected by the formation of the *EEC*. This finding, in turn, suggests that an underestimation of the *EEC* effect on member trade could result if a base year beyond 1958 is used to represent normal preintegration trade.

of the United Kingdom from the *EFTA* effect moves the estimates more in line with the *TD* hypothesis, positive values still remain for France and Italy.

REFERENCES

B. Balassa, *The Theory of Economic Integration*, Homewood 1961.

———, "Trade Creation and Trade Diversion in the European Common Market," *Econ. J.*, Mar. 1967, *77*, 1–21.

M. K. Carney, "Developments in Trading Patterns in the Common Market and EFTA," *J. Amer. Statist. Assn.*, Dec. 1970, *65*, 1455–59.

M. E. Kreinin, "Trade Creation and Diversion by the EEC and EFTA," *Economia Internazionale*, May 1969, *22*, 273–80.

H. J. Linnemann, *An Econometric Study of International Trade Flows*, Amsterdam 1966.

J. Tinbergen, *Shaping the World Economy: Suggestions for an International Economic Policy*, New York 1962.

R. Triffin, *Europe and the Money Muddle*, New Haven 1957.

E. M. Truman, "The European Economic Community: Trade Creation and Trade Diversion," *Yale Econ. Essays*, Spring 1969, *9*, 199–257.

P. J. Verdoorn and F. J. M. Meyer Zu Schlochtern, "Trade Creation and Trade Diversion in the Common Market," in *Integration Européene et Réalité Economique*, Brussels 1964.

J. Waelbroeck, "Le Commerce de la Communauté Européene avec les Pays Tiers," in *Integration Européene et Réalité Economique*, Brussels 1964.

I. Walter, *The European Common Market*, New York 1967.

European Free Trade Association, *EFTA Trade: 1969*, Geneva 1970.

———, *The Effects of EFTA on the Economies of Member States*, Geneva 1969.

International Monetary Fund, *Direction of Trade*, Washington, various annual issues, 1951–67.

———, *International Financial Statistics*, Washington, supp. 1967–68 and various issues, 1967–69.

National and Commercial Banking Group Ltd., "'The Six' and 'The Seven' in Europe," *Three Banks Rev.*, Mar. 1960, *12*, 14–32.

Organization for European Economic Cooperation (OECC), *Twelfth Annual Economic Review*, Sept. 1961.

Organization for Economic Cooperation and Development (OECD), *National Accounts of OECD Countries*, Paris, 1957–66 and 1958–67.

United Nations, *Monthly Bulletin of Statistics*, various issues, New York 1961–69.

U.S. Naval Oceanographic Office, *Distance Between Ports*, Washington 1965.

Black-White Differences in Returns to Schooling

By FINIS WELCH*

Investments in schooling in the United States have earned high returns in the aggregate, but at the same time estimated returns to blacks have been low. This paper updates previous estimates, showing that underlying the low averages reported for cross-sections of the black population there is a strong upward drift. Returns to blacks schooled in the 1920's and 1930's were so low that relative to whites, black income fell as school completion levels rose. Recent evidence shows, on the other hand, that returns, as a fraction of earnings, for blacks schooled in the 1950's and 1960's exceed returns to whites. These estimates imply that income gains that were earlier realized by whites as average educational levels increased are now being realized by blacks.

I. Earlier Estimates

In 1967, I observed that to a southern rural black, schooling was a poor investment, "A non-white with no schooling will receive 81 percent of the income of a similar white. Yet, for non-whites, school attendance increases income at a rate which is only 28 percent of the corresponding increases for whites" (p. 235). This estimate was based on the 1960 Census for earnings in 1959. Relative earnings of around 80 percent for blacks with no schooling, fall to less than 50 percent for college graduates. Using the same data, Giora Hanoch observed that " . . . internal rates of return for Non-whites are generally low and relatively erratic . . . " (p. 325), and William Landes estimated that outside the South, schooling returned only about 20 percent as much to nonwhites as to whites.

These are three of several income studies showing similarly pessimistic results, that the educational route to economic and social mobility was apparently closed to blacks. More recently this cloud of pessimism has been expanded to hover over quality of schooling as well. Christopher Jencks and his coauthors argue that family background factors dominate measures of academic achievement, leaving little room for attributes of schools to have an effect. And in any case, measured academic achievement (as distinguished from years of attendance) has little effect upon earnings. That this link between school traits and income is twice filtered through the screen of measured achievement seems not to matter. This, even though Herbert Gintis, one of Jencks' coauthors, has argued that existing achievement tests are capable of explaining only between 10 and 20 percent of the observed earning power of schooling. The view offered here is no more direct than that of Jencks. I simply compare very broad trends in school characteristics to trends in the return to schooling. It seems, on the basis of this comparison, that there may be an important relationship between

* University of California, Los Angeles and the Rand Corporation. This research is supported by a grant from the U.S. Office of Economic Opportunity to the National Bureau of Economic Research. I am indebted to Iva Maclennan, Steve Rose and Hy Sanders for valuable assistance and I am also grateful to Barry Chiswick, Victor Fuchs, Thomas Johnson, William Landes, Jacob Mincer, Melvin Reder, and Lewis Solmon for comments on an earlier draft.

schooling quality and income, or at least the possibility that such a relation exists should not be dismissed without further research.[1]

In reference to the early estimates, I think we erred by not considering the full implications of secular changes in schooling quality. The estimates are based on observations around 1960, and at that time the majority of the black adult population would have been in school in the decades of the 1920's and 1930's. During those periods, blacks attended school only about two-thirds as many days as whites. Furthermore, approximately one-third of all black students enrolled in public elementary and secondary schools in the segregated South were enrolled in the *first* grade. There were typically more than twice as many first as second graders—suggesting that retardation rates exceeded 100 percent, i.e., the average black pupil spent something more than two years in the first grade. Also, southern schools spent a little more than three times as much on white as on black pupils. In this context, it should not be surprising that schooling contributed much less to black than to white income. Given the substantial improvement in the relative quality of the schools attended by blacks in the past few decades, it should not be surprising that vis-à-vis the return to schooling, young blacks—more recent graduates—fare better in comparison to whites. And that, pure and simple, is what the data show.

II. A Reexamination

In this paper, the contributions of schooling to earnings are compared between blacks and whites for two bodies of data. The first is the 1 in 1000 sample from the 1960 Census and the second is the Survey of Economic Opportunity (*SEO*) Census income refers to earnings in 1959, and *SEO* refers to 1966 earnings. Comparison is restricted to urban males (Negro or white) who are not in the military, are not enrolled in school, and are not working without earnings. In the subsamples drawn, the Census contains 14,918 whites and 1,749 Negroes and *SEO* contains 6,379 whites and 3,442 Negroes.[2] The *SEO* data are potentially the richer source for making black-white earning comparisons because the sampling procedure insured a larger proportion of blacks. The *SEO* increased the percentage of blacks by adding to a national random sample of about 30,000 individuals, a companion sample including about 20,000 persons from "poverty" areas. It is unfortunate that stratification was based upon an income correlate, for this carries the potential of introducing biases via "creaming" of the economic successes.[3]

The point of departure for the income comparisons is the stratification of the individual observations by race and an estimate of the time of entry into the labor

[1] Lewis Solmon has recent evidence linking college traits to earnings. Also, the work of Frank Stafford and George Johnson is relevant.

[2] The Census contained observations for 56,732 males between 14 and 70 years of age who, if they reported a positive number of weeks worked, reported positive earnings. Of these, 40,065 were eliminated: 6,835 were rural; 26,902 either had "negative" work experience, were in school, or had more than 25 years experience; 169 worked without pay; and 529 were neither Negro nor white, the remainder were either in the Armed Forces, reported positive earnings for zero weeks worked, or there was missing information on working class. The *SEO* sample of males 14 to 70 years contained 28,552 persons. Of these, 18,231 were eliminated: 378 were in the armed forces; 6,000 had rural residence; 10,992 had either negative experience, were in school, or had experience in excess of 32 years; 31 worked without pay or were missing information on working class an additional 230 had race other than Negro or white; others eliminated had inconsistencies between earnings and weeks worked.

[3] As schooling increases income, it increases the probability that an individual will not be in the sample. Thus, as schooling increases, we observe increasing proportions of those who are "losers" relative to the norm, and systematically understate the increase in "expected" income associated with schooling.

force. This estimate was expressed as years of work experience.[4] The Census experience classes are 1–4, 5–12, and 13–25 years. These workers would have entered the labor force approximately between 1955–58, 1947–54, and 1934–46, respectively. Their schooling would presumably have occurred in the years prior to entry into the workforce. The *SEO* sample is conformably partitioned into experience classes of 1–3, 4–7, 8–11, 12–19, and 20–32 years. Since *SEO* refers to 1966 income, the last three classes correspond to the 1959 Census cohorts. The first two classes represent entry into the labor market between 1963–65 and 1959–62, respectively.

The estimated earnings equations are of the form:

$$(1)\quad \log Y = a_0 + a_1S + a_2 \log W + a_3ZY + a_4F + a_5SF + u$$

Observations are of individuals and variables are defined as follows:

1) *Log Y* is the natural logarithm of annual earnings when earnings are positive; it is set equal to zero when earnings are zero.

2) *S* refers to years of school completed. The coefficient estimate, $\hat{a}_1$, is thus an estimate of the proportional increase in earnings associated with an extra year of schooling (when other independent variables are held constant).

3) *Log W* is the natural logarithm of weeks worked in the year for which earnings are reported. The coefficient $\hat{a}_2$ estimates the elasticity of earnings with respect to time worked.

4) *ZY* is a dummy variable that takes the value one when earnings are zero, and is zero otherwise.[5]

5) *F* is a dummy variable that equals one if the individual is a federal employee and zero otherwise.

6) *SF* is a measure of the federal government's share of the industry in which the individual is employed. The share is equal to one for federal employees.

Thus $(\hat{a}_4+\hat{a}_5)$ estimates the percentage increase in earnings associated with direct federal employment (as opposed to working in the private nongovernmental sector), and $\hat{a}_5$ estimates the percentage increase in income associated with indirect federal employment. In a period of civil rights activism one might expect the earliest effects to appear for federal employees and in private firms in which the government has the most obvious control. These variables are included to try to identify such effects.

Regression estimates appear in Tables 6 and 7 and are summarized here.

[4] In accordance with Hanoch, I assume the following ages of entry into the labor market:

Years of School Completed	0–7	8	9–11	12	13–15	16	17+
Age at First Year Out of School	14	16	18	20	23	26	28

Experience is simply defined as current age less the estimated age at leaving school.

[5] The coefficient $\hat{a}_3$ is the negative predicted value of the average *log* of earnings for those who have no earnings. The prediction is the weighted sum of the mean levels of the other independent variables weighted by the regression coefficients. In principle, the coefficients on other independent variables are smaller, in absolute value, than they would be if the regression were restricted to those observations with positive earnings. The relationship between the coefficients reported here and those obtained in regressions restricted to positive incomes is straightforward: First, partition total variance of the ith independent variable into variance about the group (those with and without income) means and variance within each group about its mean. Let f_i represent the fraction of within-group variance attributable to those who have income. Then $\hat{a}_i = f_i\hat{b}_i$ where $\hat{a}_i$ refers to regression coefficients reported here and $\hat{b}_i$ refers to coefficients obtained from restricting the sample to persons with positive income. For the constant or intercept term, the weeks worked, and the government employment variables, a constant is assigned when earnings are zero and for these variables $f_i = 1$. Therefore the schooling coefficient is the *only* parameter estimate that is affected by introducing the zero earnings dummy.

The return to schooling is broken into two components, a direct and an indirect effect. Empirically we observe that not only do persons with more schooling receive higher wages per unit of time spent on the job but, in addition, these persons devote more time to their jobs: they work more hours per week and more weeks per year.

The direct wage effect is measured by $\hat{a}_1$. It estimates the percentage increment in annual earning associated with an extra year of schooling *holding weeks worked and the probability of zero earnings constant.* It is this component of the return to schooling which is most closely linked to productivity and therefore is most obviously affected by trends in quality of schooling. When education received in a school year is of higher quality, the coefficient is expected to be higher.[6] This coefficient (plus one) is the relative wage of a person with an extra year of schooling, and can be taken in a competitive market as an estimate of the marginal rate of substitution of a worker with given schooling for a worker with an extra year. As such, it is sensitive to factor proportions and in the aggregate may fall as average schooling levels rise.

The second component in measuring the earnings gain from schooling is the compounding of earning differentials that result from comovement between wage rates and time devoted to work. Here the indirect effect refers to percentage increments in earnings related to the increased number of weeks worked and the reduced probability of having zero earnings that are associated with increased schooling. This effect is measured as:

$$\hat{a}_2 \left.\frac{d\,log\,W}{dS}\right|_E + \hat{a}_3 \left.\frac{dZY}{dS}\right|_E \tag{2}$$

The terms

$$\left.\frac{d\,log\,W}{dS}\right|_E \quad \text{and} \quad \left.\frac{dZY}{dS}\right|_E$$

are auxiliary regression coefficients on schooling holding experience constant. Two auxiliary regressions are estimated, one in which the dependent variable is *log W* so that the coefficient is the proportional increase in weeks worked associated with an added year of schooling. In the other the dependent variable is the *ZY* dummy and the schooling coefficient is an estimate of the change in the probability of having zero earnings associated with an extra year of schooling. These coefficients are reported in Tables 6 and 7 with estimates of the earnings equation. For computing variances of estimates, these auxiliary coefficients are treated as constant.

A complete specification of factors determining the indirect return of schooling is beyond the scope of this paper, but I will note a few of the more obvious possibilities. First, if unemployment is involuntary, the auxiliary relation between time worked and schooling may refer to statistical screening of employees. Schooling reduces job search time. Insofar as this phenomenon is important, we expect a contracyclical component in the indirect return. In a period like 1959 of relatively high unemployment, the impact of involuntary unemployment is greatest for the less schooled. We also expect this auxiliary correlation between schooling and time worked to attenuate over the working career as better measures of worker quality emerge. The second alternative is that the auxiliary relation between work and schooling (the indirect effect) is simply a labor supply response to the direct or wage rate effect (see C. M. Lindsay).

[6] As an example, define education as the product, quantity times quality of schooling. In the model, $log\ W=\alpha_0+\alpha_1 E$, where W is the hourly wage and E is education. Since $E=QS$, quality times quantity of schooling, the percentage return to quantity is $\alpha_1 Q$.

TABLE 1—PROPORTIONAL INCREMENT IN INCOME ASSOCIATED WITH ONE ADDITIONAL YEAR OF SCHOOLING[a]

Experience Class	White			Black		
	Total	Direct	Indirect	Total	Direct	Indirect
Census (1959)						
1–4	.208	.095	.111	.317	.079	.231
5–12	.134	.081	.052	.149	.065	.077
13–25	.099	.064	.035	.059	.040	.015
SEO (1966)						
1–3	.306	.146	.157	.519	.230	.289
4–7	.191	.106	.084	.256	.142	.111
8–11	.149	.082	.067	.192	.079	.109
12–19	.143	.090	.053	.130	.062	.065
20–32	.110	.069	.041	.096	.049	.042

[a] Standard Errors are not reported, but were typically small relative to the coefficients. For total returns, t-ratios ranged between 5.2 and 33.0, between 14.6 and 32.0 for direct returns, and for indirect returns, between 2.3 and 11.7 for whites and 1.8 to 5.0 for blacks.

If the supply of labor is positively sloped, we then expect the direct effect, which captures the wage increment, to be positively related to the indirect effect, which embodies the resultant work response.[7]

Table 1 reports estimates of the total, direct, and indirect returns to schooling. The sum of the direct and indirect effects is not constrained to equal the total effect, because a provision for federal employment is also included. The covariances between federal employment and schooling are consistently positive, both for blacks and whites, but the contribution to the return to schooling is numerically trivial even though it occasionally is "statistically significant."

Several patterns in these coefficients reported in Table 1 are of interest. The most obvious is the attenuation in the cross-section of both direct and indirect returns as experience increases or as "vintage" decreases. In 1959, a year of fairly high unemployment, the indirect effects actually exceed the direct effects at lower levels of work experience, although the pattern quickly reverses itself as experience increases. Further, the share of the indirect in the total effect is larger for blacks than for whites. This adds support to the hypothesis of involuntary unemployment, with the heavier burden falling upon blacks and upon persons with little experience, i.e., persons for whom the years of school completed represents an important part of the information available for assessing productivity.

In the 1959 data, the evidence is that for persons with 1–4 years of experience, black earnings rise relative to white earnings as school completion levels increase. This point has not been previously noted. For persons with 5–12 years of experience, the black/white earning ratio is insensitive to schooling and for persons with 13–25 years of experience relative earnings

[7] In fact, if there is no involuntary unemployment and if schooling is competitively purchased such that there is no income or wealth effect, the ratio of the indirect to the direct effect is an estimate of the substitution component in the labor supply function. The indirect effect is the time worked response to the wage change accompanying increased schooling so that the ratio of this to the wage response to schooling gives an estimate of the rate time at work responds to wage increments.

of blacks fall as schooling increases.

The 1966 data presents patterns that are generally similar to those found for 1959. At low levels of experience, the black/white earning ratio rises with schooling, but at higher levels of experience, the opposite is true. Interestingly enough the share of the indirect in the total return to schooling is similar between blacks and whites in the relatively tight labor market that prevailed in 1966.

Because the indirect effect is likely to contain important life cycle and business cycle influences, a comparison of the effects of secular trends in the relative quality of black schooling is more safely founded in comparisons of direct effects alone.

The first panel of Table 2 presents t-statistics for black-white differences in the direct return to schooling in 1959 and 1966 within experience classes. Panel II gives t-statistics for the black and white cohorts for differences in direct returns between 1959 and 1966. These statistics simply confirm what is obvious in Table 1. For those with 1–4 years of experience in 1959 or 8–11 years in 1966 there is no significant black-white difference. As experience extends to encompass persons who entered the labor force before the mid-1950's, the difference cumulates becoming "significantly" higher for whites especially in the 13–25 and 20–32 experience classes. But, for persons entering the workforce in the 1950's the evidence is that returns are significantly greater for blacks. That is, viewed as a proportional contribution to earnings, schooling yields more to blacks than whites when comparisons are restricted to recent vintages. The earlier vintage comparisons yield the opposite conclusion.

That this effect is primarily one of vintage—not a life cycle effect in which schooling initially offers more to blacks but quickly peters out—is evidenced in the cohort comparisons. Of six comparisons, the direct returns to schooling are remarkably similar in a numeric sense, and statistically they change significantly in only one case of six as the cohorts add seven years of experience. More generally, the

TABLE 2—STUDENT t-STATISTICS FOR COMPARISONS OF DIFFERENCES IN DIRECT RETURNS TO SCHOOLING

I. Black-white differences (coefficient difference relative to standard error).

Census (1959)		*SEO* (1966)	
—		1–3	+11.83
—		4–7	+2.37
1–4	−0.69	8–11	−0.39
5–12	−1.74	12–19	−3.47
13–25	−3.81	20–32	−3.45

II. Cohort differences 1959–1966.

Approximate Year of Entry into Labor Force	Black: Direct Schooling Coefficient 1959	Black: Direct Schooling Coefficient 1966	Black: t-ratio on Coefficient Difference	White: Direct Schooling Coefficient 1959	White: Direct Schooling Coefficient 1966	White: t-ratio on Coefficient Difference
1955–58	.079	.078	.04	.095	.082	+1.96
1947–54	.065	.062	.27	.081	.090	−1.86
1934–46	.040	.049	−1.15	.064	.069	−1.39

twelve equations in which the six cohort comparisons are based were "stacked" into a single regression equation for which it is possible to jointly impose the constraints of cohort equality for the schooling coefficients in 1959 and 1966. The computed $F(6;24,705)$-statistic on this constraint is 1.73 with a 0.05 critical level of 2.10.

This is not to say that there are no life cycle effects associated with the return to schooling,[8] but if such effects exist they are swamped in these data by the general "tightening" of the labor market that occurred between 1959 and 1966. To summarize, the evidence is that, viewed as a proportional contribution to earnings, returns to schooling have increased for blacks relative to whites. For those entering the workforce in the 1930's and early 1940's (who were schooled over the ten to fifteen years preceding entry to the market), returns for blacks were significantly less than for whites. Yet the opposite is true for those entering the labor market in the 1960's. Furthermore, in the cohort comparison this return does not attenuate as experience increases. Since rapid attenuation occurs in each cross-section, the implication is that vintage is an important explanation. This interpretation is also presented by Leonard Weiss and Jeffrey G. Williamson in a separate analysis of the *SEO* data.[9]

This vintage effect could be interpreted as changes in labor market conditions with changing attitudes toward discrimination. Suppose that the labor market is such that a person's career profile is largely determined by conditions that exist at the time he enters the workforce, and that "taste for discrimination" has been falling over time. This would predict the improving relative position of blacks, but would not explain the obvious vintage effect for whites. As an alternative, consider the effects of quality of schooling. The evidence is that real inputs into predominantly white schools have increased through time and that black schools have improved relative to white schools. This is sufficient to explain the observations reported in Tables 1 and 2. Given this possibility, turn to evidence of schooling's quality.

III. Schools for Blacks

Until we are more knowledgeable about the technology of learning and about exactly what it is that schools produce that allow more schooled persons to earn more, comparisons of schooling quality will be superficial at best. Nonetheless, we do have fragmentary evidence that both in terms of the facilities schools offer to students and in terms of measured achievement, blacks are gaining relative to whites. This is not to say that we have achieved parity, but only that trends are in the "right" direction.

[8] In fact, at very low levels of work experience there is good reason to expect a life cycle component in the direct return as well as in the indirect return. Bear in mind that an extra year of schooling, experience constant, implies that age also increases by one year. In fact, in the estimates used here, the age of entry into the labor market rises something more than a year with an added year of schooling. At relatively young ages this maturation effect, included in the return to schooling, may be important.

[9] It is not clear that a distinction between life cycle and vintage effects is possible using a single cross-section. Weiss and Williamson observed that over the age interval, 20–49 years, and within the schooling interval, 8–12 years, the percentage increment in earnings from schooling falls as age increases. This is not true for persons who have attended college nor is it true of those who are 50 years old or over. Weiss and Williamson purport to test the impact of inferior southern schooling by using the variable for residence at age 16 that is available in *SEO*. Unfortunately, they did not allow interaction between schooling and this early residence variable, so that the marginal value of schooling is assumed independent of location. This formulation is then not a test of quality. In fact, given the evidence of trends through time presented here, I would argue that the appropriate test concerning quality of schooling involves interaction between vintage, early residence, *and* schooling. The introduction of dummy variables for early residence does not adequately specify the role of schooling's quality.

In Table 3, summary statistics are offered for segregated southern schools for selected years between 1919–20 and 1953–54. With the Supreme Court decision of 1954, the publication by the federal government of segregated school system statistics stopped. It is difficult to say how seriously these comparisons should be taken. Clearly, not all blacks attended school in the South. Less clearly, the relative quality of school offerings may have been better outside the South. Nonetheless, the data of Table 3 indicate a very strong upward trend in virtually every measure of schooling quality I could find, and it is this trend I want to stress.

The change that may have been of greatest importance, in terms of total learning in this century, has been the convergence in the length of school terms between urban and rural areas and between the South and other parts of the country. Because blacks were concentrated in the rural South prior to World War II, they may have been penalized most by the shorter school terms, and may have gained most by the general trend to equality of the term. As is indicated in the third row of Panel I, in 1919–20, blacks attended school only two-thirds as many days each year as did whites. Differences in days attended were more the result of

TABLE 3—COMPARISON OF TRENDS IN ENROLLMENT, ATTENDANCE, AND EXPENDITURES FOR SEGREGATED NEGRO PUBLIC ELEMENTARY AND SECONDARY SCHOOLS IN THE SOUTHERN UNITED STATES TO TRENDS FOR ALL SCHOOLS, 1919–54

Characteristic:	1919–20	1929–30	1939–40	1949–50	1953–54
I. School Term and Attendance					
1. Average number of days schools were in session					
a) Negro Schools	119	132	156	173	177
b) All schools	162	173	175	178	179
2. Percent of pupils enrolled in average daily attendance					
a) Negro Schools	67.3	72.1	80.4	85.3	85.1
b) All Schools	74.8	82.8	86.7	88.7	88.9
3. Average days attended per pupil enrolled					
a) Negro Schools	80	97	125	148	151
b) All Schools	121	143	152	158	159
II. Enrollment and Grade Level					
4. Percent of pupils in public elementary and secondary school who are enrolled in the first grade					
a) Negro Schools	36.8	34.4	26.0	19.5	16.5
b) All Schools	22.9	16.2	11.9	12.6	12.7
5. Percent of pupils in public elementary and secondary schools who are enrolled in secondary school					
a) Negro Schools	1.6	4.9	10.5	14.1	16.2
b) All Schools	10.2	16.1	26.0	22.7	21.8
III. Salaries and Instructional Expense					
6. Average salary per member of instructional staff					
a) Negro Schools	—	—	$ 601	$2143	$2861
b) Southern White Schools	—	—	1046	—	3384
c) All Schools	$871	1420	1441	3010	3825
7. Current expenditure per pupil in average daily attendance					
a) Negro Schools	—	15[a]	19	—	110[b]
b) Southern White Schools	—	49[a]	59	—	181[b]
c) All Schools	—	87	88	209	265

Source: Various issues of Biennial Survey of Education, David Blose.
[a] Refers to 1931–32 instead of 1929–30.
[b] Instructional expenses only.

differences in the length of the school term than in attendance rates. By 1953–54, black-white differences in days attended had virtually disappeared. The overwhelming difference between black and white schools in the 1920's and 1930's was in the composition of enrollment. In this period, typically one-third of all black elementary and secondary school pupils were enrolled in the *first* grade. Retardation rates appear to have been phenomenal in both the first years of elementary and secondary school. For example, prior to 1940 there were consistently more than twice as many black pupils enrolled in the first as in the second grade, and about one and one-half as many in the first as in the second year of high school. Since school enrollment was increasing at only 1 or 2 percent per year, the bulk of this discrepancy is retardation between grades one and two, and in high school it is a combination of retardation and "dropping out."[10]

By 1953–54, the ratio of first to second graders had fallen to 1.45 and the number in the first year of high school, relative to the second, had fallen to 1.30. Another important trend in the composition of enrollment was the rapid increase in the proportion attending secondary school. In 1919–20, only 1.6 percent of all black pupils in public schools were enrolled in high schools (in comparison to 10.2 percent of white pupils). By 1953–54 this proportion increases to 16.2 percent.

Expenditures on the black schools appear to have lagged behind the increases in attendance, enrollment, and the reductions in retardataion, but the trend bebetween 1940 and 1954 is impressive. Teacher salaries increased from a relative level of 57.5 percent of the salaries of southern white teachers to 85.0 percent in the later year. Similarly, expenditures rose from 31 percent of southern white levels in 1931–32 to 61 percent in 1953–54. Perhaps the levels of relative expenditures were a social crime at the time of the segregation decision, but the trend had been upward and had apparently accelerated after the beginning of World War II. And it is this upward trend that is an outstanding candidate for explaining the substantial trend in the value of schooling to blacks in comparison to whites. Recall that those in the experience class 13–25 years in 1959 would have last attended school between 1924 and 1945. Maybe the question of why returns were so low should have been reversed: If expenditures by school systems are the norm, then the relative value of black schooling is only slightly less than expected.

There is scattered evidence that the southern story of separate but unequal schools was partly true of the North—slightly less separate and slightly less unequal. Harold Baron offered some evidence for Chicago in 1961, 1963, and 1966. He argues that in 1961 there were significant differences in expenditures within the city's educational system which favored white students. By 1966, this expenditure difference had disappeared.

The Board of Education in New York City annually gives nationally standardized reading tests to third, sixth, and eighth grade pupils, and school average scores are available. Table 4 summarizes estimates of black-white differentials for selected years between 1957 and 1969. The technique used for measuring these differences is based upon the identification of *de facto* segregated schools, since only school averages, not individual scores, are available. Once again, the story of reading score differentials is pretty much the same as for income differences and for the segregated southern schools of an earlier period.

Viewed in absolute terms, there are very

[10] Since there were always more students enrolled in the first year of high school than in the last year of grade school, the evidence is for retardation.

TABLE 4—COMPARISON OF BLACK-WHITE READING SCORE DIFFERENTIALS BY GRADE LEVEL; *De Facto* SEGREGATED NEW YORK CITY SCHOOLS, SELECTED YEARS, 1957–69[a]

	Grade in Which Test was Administered	
	Third	Sixth
Norm (National Average)	3.5	6.5
New York City Schools:		
1957		
Black	2.67	4.88
White	4.07	7.81
1960		
Black	2.87	5.22
White	4.14	8.10
1965		
Black	3.19	5.67
White	4.53	8.13
1969		
Black	3.31	5.38
White	4.37	7.60

Source: The Board of Education of New York City, unpublished data.

[a] These data are weighted averages of school reading score averages within each class. Third and sixth grade scores refer to elementary schools. The scores are normalized to allow for year-to-year variation in the time the test is administered. The criterion for selecting schools for these computations varied from year-to-year as data availability permitted.

Black Schools: The object was to select a modest number of reasonably large schools with the largest proportions of black students. The third grade data refer to from 21 to 28 schools, each having 700 or more pupils with at least 90 percent black. The sixth grade data are for the same schools, except that in 1965 and 1969 only eleven of these elementary schools had sixth grades.

White Schools: The number of elementary schools ranged between 23 and 27 with enrollment exceeding 800, at least 97 percent white.

large black-white differentials, and these differences compound as years completed increase. Whatever forces result in differences by the third grade appear to continue to operate at the sixth grade. The important evidence of Table 4 lies in the trend between 1957 and 1969. For third graders, a 1.40 years reading deficit eroded to 1.06. In twelve years, black third graders in New York City improved an average of .64 years in reading achievement This is reason for optimism. The reduction in the deficit for the sixth grade is a little more ambiguous. The deficit falls from 2.93 years to 2.22—a larger absolute improvement, but about the same fraction (25 percent) of the original deficit as for third graders. Certainly the gain for blacks from 4.88 to 5.38 is important, but the norm is 6.5, and the deficit remains large. Notice that children in the sixth grade in the spring of 1969 would have entered school no later than fall of 1963, and would therefore have missed the Headstart Programs. Also, given the rapid rise in black school completion levels, the average education levels of parents of black third graders in 1969 would have exceeded the average schooling of parents of black sixth graders so that these increments may refer more to changes in home than in school environment. For purposes of this analysis the important point is that the "knowledge" associated with given school completion is rising for blacks relative to whites. The question of why these changes occur is not relevant.

The fragments of evidence tell a consistent story. Schools for blacks may be inferior to those for whites and much improvement is still in order, but the record is undeniable. Improvement has occurred.

IV. Earnings and Returns to Schooling

Table 5 reports average earnings and the associated earning increment from a single year of schooling, estimated from the earning equation when schooling is held constant at 12 years.

From 1959 to 1966 the black/white earnings ratio increased by 0.12 and 0.13 for the two less experienced cohorts and by 0.23 in the third cohort. This suggests that much of the gain over this seven-year period may be serendipity of the business cycle, subject to erosion as labor mar-

TABLE 5—ESTIMATED AVERAGE ANNUAL EARNINGS AND THE VALUE OF AN EXTRA YEAR OF SCHOOLING BY RACE AND EXPERIENCE[a]

(Schooling = 12 years)

Experience Class:	Annual Earnings			Return to Schooling (direct component only)		
	Black	White	Black/White	Black	White	Black/White
Census (1959)						
1–4	$1500	$2620	.57	$120	$250	.41
5–12	2510	4700	.53	160	380	.43
13–25	2360	5830	.41	90	370	.25
SEO (1969)						
1–3	2830	3160	.90	650	460	1.41
4–7	3880	5024	.77	550	530	1.03
8–11	4160	5968	.70	330	490	.67
12–19	4420	6800	.65	270	610	.45
20–32	4300	6690	.64	210	460	.46

[a] Average income is predicted from the regression equations of Tables 6 and 7. Where an auxiliary structure is estimated between schooling, S, and another independent variable, x, the mean level of that variable is adjusted to $S=12$, according to $\bar{x}(s-12)=\bar{x}+b(12-\bar{s})$, where b is the auxiliary regression coefficient. Mean income is estimated as the antilog of the predicted mean *log* of earnings plus one-half of the residual variance. The variance adjustment results from an assumption that the residuals are lognormal and increases the income predictions by 15 to 30 percent. The estimated return to schooling is the schooling coefficient (the direct return) times estimated mean income at 12 years of schooling.

ket conditions weaken. It seems reasonable to predict that had the 1966 observations occurred alongside the 1959 labor market, the earning ratios would decline to their 1959 levels. If so, then instead of relative earnings between .77 and .90 for those beginning work in the 1960's, the ratios would have been something like .64–.77. Although vintage effects appear to be pushing toward earnings parity, the gap remains large.

Variables for direct and indirect governmental employment were added to the regression equations to try to disentangle the effects of activities of the federal government on relative black earnings. The results are that, although these variables are statistically significant, the total effect, allowing for employment shares, is numerically small. Direct governmental employment increased average black earnings relative to the white average by 1 to 2 percent in both 1959 and 1966, and the effects of indirect governmental employment are similar. Blacks who work for the government (directly or indirectly) gain relative to whites, but the proportions so employed are small enough that the total effect remains small. Nonetheless, since the earning differential between private and governmental employment is large for blacks, there is an apparent incentive to acquire these jobs.

V. Summary

This paper is an empirical attempt to identify some of the structural components of the observed rapid rise in the relative earnings of blacks in the United States between 1959 and 1966. It is clear that much of the gain is associated with tightening labor markets: In 1959 white high school graduates in all experience classes worked 15 to 20 percent more "weeks" than blacks whereas the discrepancy fell to about 2 percent in 1966.

But, even though this cyclic component is important in accounting for the rise in relative earnings, there is also a vintage effect to be considered. In comparison to whites, blacks who have entered the workforce in recent periods fare better than earlier entrants.

Part of this gain results from a convergence of school completion levels. For example in the Census, blacks with 1–4 years of experience averaged 11.1 years of schooling as compared to a white average of 12.8. Yet for those with 13–25 years of experience, the black average was 8.8 years and 11.4 for whites. Clearly part of the gain in relative earnings of younger blacks comes from this relative increase in average school completion, but a more important part of the gain comes from the combined effect of increased schooling and an increased return to schooling. For example, based upon the direct returns to schooling (holding constant weeks worked and the probability of zero earnings) for the cohort with 13–25 years work experience in 1959, the rise in black average school completion would have resulted in an increase in earnings of 9.2 percent, and taking the increase in white schooling into account, relative earnings would have increased by only 0.2 percent. Yet based upon the direct returns realized for the cohort with 1–4 years of experience, the increase in black earnings resulting from increased average schooling would have been 18.2 percent and earnings, relative to whites, would have risen by 5.0 percent.

Clearly, the complete picture of relative black earnings cannot be obtained from a reference to quality of schooling, but examination of the data suggests that schooling quality is important. For both blacks and whites, the cross-sectional evidence is that the proportional contribution of schooling to annual earnings declines sharply as experience increases, or, since these are cross-sections, as vintage decreases. Yet in comparing six cohorts (three experience classes for blacks and whites separately) this attenuation does not occur in the seven-year interval between 1959 and 1966. The direct return to schooling is remarkably constant over this period, suggesting that the experience-vintage effect of the cross-section is really only a vintage effect: More recent entrants into the workforce enjoy higher returns to schooling than do those persons schooled in an earlier period. Further, this upward trend is stronger for blacks, such that the percentage contribution of schooling to earnings is larger than for whites in comparing persons who entered the labor force in the 1960's but is significantly lower for those who entered in 1930's and 1940's.

To get a feel for the underpinnings of this change, I turn to a comparison of black-white differences in statistics of schools. There, the data are clear: Through time the relative quality of black schooling has risen rapidly. In the segregated South there were important differences in enrollment, attendance, and expenditure patterns at the time of the 1954 desegregation decision, but these differences are dwarfed in comparison to those of the 1920's and 1930's. Since 1954, partial evidence of continuing convergence in schooling quality is offered by referring to a study of the Chicago schools that found black-white expenditure differences vanishing by 1966. For New York City schools, reading score data for *de facto* segregated schools indicate that about one-third of the discrepancies existing in 1957 had disappeared by 1969.

There are, of course, many alternative explanations for the vintage effects of rising relative black income: a downward drift in market discrimination being the simplest and most obvious. It seems reasonable to expect that whatever social, political, and judicial forces allowed convergence in offerings to black and white students also have allowed convergence in

offerings to black and white workers. The role of alternative hypothesis is confounded because most would not be exclusive of the quality phenomenon, but would be supplementary to it. It would be helpful to be able to identify the partial effects of the several forces that have been operative, but I doubt that available data are capable of such fine distinction. Certainly the data reported here are not.

In behalf of the quality of schooling hypothesis let me summarize the trends revealed here with which alternatives must contend. First, not only have relative black incomes increased but the gain has been greatest for higher school completion levels. Second the phenomena of rising returns to schooling is not only true in comparing black relative to white incomes but holds within the races as well: Young blacks fare better in comparison to young whites than do older blacks in comparison to older whites *and* schooling contains more of an income boost for young blacks and young whites than for older generations of their own races.

TABLE 6—REGRESSION RESULTS

(Except in auxiliary regressions, the dependent variable is the natural *log* of annual earnings)[c]

	Census 1959: Black			*SEO* 1966: Black				
Experience Class	1–4	5–12	13–25	1–3	4–7	8–11	12–19	20–32
Years of Schooling	.079	.065	.040	.230	.142	.078	.062	.049
	(.023)	(.009)	(.006)	(.020)	(.014)	(.009)	(.007)	(.005)
Log of Weeks Worked	1.158	1.335	1.157	.910	.833	1.053	1.083	1.305
	(.097)	(.069)	(.055)	(.103)	(.119)	(.127)	(.127)	(.078)
	.143	.040	.009	.090	.024	.027	.016	.012
Zero Earnings	−3.039	−2.635	−3.556	−5.467	−6.227	−6.101	−6.116	−5.740
	(.387)	(.289)	(.218)	(.234)	(.279)	(.268)	(.264)	(.163)
	−.021	−.009	−.001	−.038	−.015	−.013	−.008	.005
Federal Employee	−1.653	−.509	−.481	−.638	−.031	−.451	−.508	−.651
	(.786)	(.316)	(.227)	(.525)	(.352)	(.206)	(.194)	(.157)
	.019	.017	.013	.007	.015	.022	.015	.016
Federal Share of Industry[a]	1.804	.793	.823	.326	.161	.628	.758	.959
	(.797)	(.316)	(.225)	(.506)	(.332)	(.198)	(.186)	(.158)
	.022	.019	.013	.015	.017	.024	.015	.016
Experience[b]	.133	.185	.028	.683	.031	−.581	−.090	.012
	(.263)	(.108)	(.055)	(.390)	(.369)	(.449)	(.141)	(.064)
Experience Squared	−.018	−.009	−.001	−.156	−.001	.032	.003	−.000
	(.051)	(.006)	(.001)	(.097)	(.034)	(.024)	(.005)	(.001)
Intercept Residual Variance	2.040	1.119	2.903	2.944	4.829	8.067	6.203	5.146
	.531	.420	.327	.676	.487	.246	.339	.337
R^2	.895	.882	.924	.891	.853	.916	.891	.913

[a] Government purchases in industry relative to *GNP* originating in industry.
[b] Experience is operationally defined in fn 4.
[c] For each variable, either 2 or 3 numbers are reported. The first is the regression coefficient, the second (in parentheses) is the standard error. When a third number appears, it is a coefficient from an auxiliary regression of this variable on schooling, other independent variables held constant.

TABLE 7—REGRESSION RESULTS

(Except in auxiliary regressions, the dependent variable is the natural *log* of annual earnings)[c]

	Census 1959: White			*SEO* 1966: White				
Experience Class	1–4	5–12	13–25	1–3	4–7	8–11	12–19	20–32
Years of Schooling	.095	.081	.064	.146	.106	.082	.090	.069
	(.004)	(.002)	(.002)	(.010)	(.006)	(.005)	(.004)	(.003)
Log of Weeks Worked	1.314	1.299	1.395	.868	.982	.693	1.149	1.256
	(.034)	(.030)	(.024)	(.073)	(.085)	(.080)	(.071)	(.058)
	.066	.030	.020	.059	.026	.021	.015	.010
Zero Earnings	−2.858	−3.252	−3.141	−6.019	−6.275	−7.151	−6.411	−6.338
	(.149)	(.130)	(.103)	(.208)	(.206)	(.191)	(.161)	(.124)
	− .008	− .004	− .002	− .017	− .009	− .007	− .006	− .004
Federal Employee	− .618	− .323	− .316	− .422	− .462	− .231	− .247	− .324
	(.117)	(.065)	(.050)	(.272)	(.160)	(.137)	(.100)	(.081)
	.004	.003	.005	.003	.006	.007	.003	.005
Federal Share of Industry[a]	.465	.231	.245	.624	.459	.331	.236	.323
	(.096)	(.054)	(.045)	(.230)	(.135)	(.109)	(.087)	(.073)
	.008	.005	.005	.006	.007	.007	.006	.006
Experience[b]	.009	.073	.029	.483	.154	− .502	.075	.027
	(.066)	(.028)	(.018)	(.268)	(.198)	(.301)	(.082)	(.042)
Experience Squared	.008	− .002	− .001	− .085	− .011	.029	− .002	− .001
	(.013)	(.002)	(.001)	(.066)	(.018)	(.016)	(.003)	(.001)
Intercept	1.825	2.045	2.160	4.284	4.919	8.559	4.939	5.405
Residual Variance	.328	.256	.270	.484	.282	.215	.246	.293
R^2	.866	.833	.835	.859	.850	.876	.877	.896

[a] Government purchases in industry relative to *GNP* originating in industry.
[b] Experience is operationally defined in fn 4.
[c] For each variable, either 2 or 3 numbers are reported. The first is the regression coefficient, the second (in parentheses) is the standard error. When a third number appears, it is a coefficient from an auxiliary regression of this variable on schooling, other independent variables held constant.

REFERENCES

H. M. Baron, "Race and Status in Schools Spending: Chicago 1961–1966," *J. Hum. Resources*, winter 1971, *6*, 3–24.

D. Blose, "Statistics of the Education of Negroes," U.S. Office of Education, circ. no. 215, June 1943.

H. Gintis, "Education, Technology, and the Characteristics of Worker Productivity," *Amer. Econ. Rev. Proc.*, May 1971, *61*, 266–79.

G. Hanoch, "An Economic Analysis of Earnings and Schooling," *J. Hum. Resources*, summer 1967, *2*, 310–29.

C. Jencks et al., *Inequality: A Reassessment of the Effect of Family and Schooling for America*, New York 1972.

W. Landes, "The Economics of Fair Employment Laws," *J. Polit. Econ.*, July/Aug. 1968, *76*, 507–52.

C. M. Lindsay, "Measuring Human Capital Returns," *J. Polit. Econ.*, Nov./Dec. 1971, *79*, 1195–1215.

L. Solmon, "The Definition and Impact of

College Quality," unpublished paper, Nat. Bur. Econ. Res., June 1972.

F. Stafford and G. Johnson, "Social Returns to Quantity and Quality of Schooling," unpublished paper, Univ. Michigan, Dec. 1970.

L. Weiss and J. G. Williamson, "Black Education, Earnings, and Interregional Migration: Some New Evidence," *Amer. Econ. Rev.*, June 1972, *62*, 372–83.

F. Welch, "Labor-Market Discrimination: An Interpretation of Income Differences in the Rural South," *J. Polit. Econ.*, June 1967, *75*, 225–40.

U.S. Department of Labor, *Manpower Report of the President*, Apr. 1971.

———, Bureau of Labor Statistics, *Patterns of U.S. Economic Growth*, Bull. 1672.

U.S. Office of Education, *Biennial Survey of Education in the United States*, "Statistics of State School Systems," various issues. (Earlier editions are by the Bureau of Education in Department of the Interior).

The Friedman-Meiselman CMC Paper: New Evidence on an Old Controversy

By William Poole and Elinda B. F. Kornblith*

In 1963 Milton Friedman and David Meiselman published a paper for the Commission on Money and Credit, in which they attempt to test the relative predictive power of Keynesian and monetary theories of business fluctuations. Their methods and conclusions were sharply disputed in papers by Donald Hester (1964a), Albert Ando and Franco Modigliani (1965a), and Michael DePrano and Thomas Mayer (1965a).[1] Our purpose is to review this controversy. Our method is the simple one of first replicating the studies over the original sample periods and then examining the postsample performance of the competing equations.

On a sociological level there can be little doubt that this controversy was, and continues to be, enormously valuable. Many economists have been stimulated to work on the questions disputed by the monetarist and Keynesian combatants, and it is probably correct to argue that two models of the *U.S.* economy—the St. Louis Federal Reserve and the FRB-MIT-PENN models—are a direct outgrowth of the controversy. What has been missing from this continuing controversy, however, is an *ex post* evaluation of the earlier empirical work. In Section I we present an evaluation of the equations offered by F-M, H, A-M, and D-M. This section also includes an evaluation of the reduced-form model estimated by Leonall Andersen and Jerry Jordan (A-J). In Section II we appraise the significance of the results.

I. Evaluation of Equations

The purpose of this section is to measure the performance of the equations offered by F-M, H, A-M, and D-M in the 1959–70 period, which is beyond the period originally used for estimation. As a check on the postsample results, we have also looked at the A-J equation. Although F-M emphasized the importance of using the longest possible historical period to test the rival theories, the subsequent papers by H, A-M, and D-M relied primarily on annual data from the 1929–58 period.[2] The data have been revised since the original studies were published, and we first reestimated all of the equations over 1929–58 period using data available in late 1971.[3] Our esti-

* Senior economist, Board of Governors of the Federal Reserve System, and graduate student, University of Rochester, respectively. Barbara McFadden provided programming assistance. We are, of course, solely responsible for the analysis and conclusions of this paper.

[1] The following abbreviations will be used in the remainder of this paper—F-M: Friedman-Meiselman; H: Hester; A-M: Ando-Modigliani; D-M: DePrano-Mayer; A-J: Andersen-Jordan.

[2] We have concentrated on equations using synchronous variables and annual data in current dollars since the original controversy centered on these equations. However, there is a brief discussion of the A-J equation which uses quarterly data.

[3] The source of national income accounts data for 1929–63 is *The National Income and Product Accounts of the United States, 1929–1965, Statistical Table*; the sources for 1964–70 are the July issues of the *Surv. Curr. Bus.* for 1968, 1969, 1970, and 1971. A few minor items not reported in these sources were obtained from the Department of Commerce. The source of money supply data for 1947–70 is "Revision of the Money Stock," *Fed. Res. Bull.*, Dec. 1970, pp. 895–98; the source of the money supply data for 1929–46 is Friedman and Schwartz, Table 1A, pp. 704–33, linked to the Federal Reserve series by an analysis of the two series over the 1947–60 period during which they overlap.

mates are very similar to those reported in the original papers.[4]

We have used a definition of M_2 that includes large certificates of deposits (*CD*) in spite of the fact that the definition in general use today excludes large *CD*'s. The reason for this choice is that neither F-M nor any of the others argued at the time of the original controversy that large *CD*'s should be excluded. At that time the argument was over the appropriate definition of the autonomous expenditures variable. Nor was there any discussion that the empirical definitions of any of the variables should be revised over time according to considerations spelled out *ex ante*. It would be impossible to perform an unbiased postsample test of the competing regression equations were we to permit changes in the definitions of the variables or in the form of the regressions. In fact, it turns out that excluding large *CD*'s from the definition of M_2 would reduce the postsample predictive performance of the M_2 equation offered by F-M by increasing the amount of underprediction.

In addition to estimates using M_2 as the definition of the money stock, we have added estimates using M_1 since the rapid growth of large *CD*'s since 1959 may have changed the significance of M_2 (defined either including or excluding large *CD*'s). The growth of *CD*'s affects the postestimation performance of the equation using M_2, but not the parameter estimates of the equation since *CD* growth occurred after the 1929–58 period of estimation. We have not examined an M_2 definition excluding large *CD*'s.

A. *Regression Results Comparing Money and Autonomous Expenditures*

By our count, twelve directly competing regressions using annual data were offered by F-M, H, A-M, and D-M to test the Keynesian and quantity theories, and we have added one more using the M_1 definition of money as a part of our evaluation of the controversy. Except for one multiple regression using the variables XA and ZA (defined below), all of the regressions use one independent variable. In every case, consumption expenditures in current dollars is the dependent variable. The definitions of the independent variables are as follows:[5]

M_1 = Currency plus demand deposits adjusted at all commercial banks

M_2 = Currency plus demand deposits adjusted plus time deposits at all commercial banks

A = Net private investment plus total government deficit on income and product account plus net foreign balance

ZA = Net private investment other than in inventories plus total government purchases plus exports

XA = Net interest paid by government minus property tax portion of indirect business taxes plus government transfers other than unemployment benefits plus subsidies less current surplus of government enterprises minus statistical discrepancy minus excess of wage accruals over disbursements

A^* = ZA plus XA

A^{**} = Net private investment other than in inventories plus federal government purchases plus exports

A^{***} = Gross private investment other than in inventories plus federal government purchases plus exports

L = Net private investment plus total government purchases plus net exports

[4] Hester did not report regression coefficients in his paper and so we have not been able to compare our estimates with his.

[5] The definitions in this list are used by the following authors: Friedman-Meiselman, M_2, A; Ando-Modigliani, ZA, XA, A^*; DePrano-Mayer, A^{**}, A^{***}; and Hester: L, L', L'', L'''.

L' = Gross private investment plus total government purchases plus exports

L'' = Gross private investment plus total government purchases plus net exports

L''' = Gross private investment other than in inventories plus total government purchases plus exports.

The postsample performance of the regressions (excluding the A-J equation which is discussed later) is analyzed in Table 1. In the table column headings, Ac refers to the actual, and P the predicted level of consumption, S is the standard deviation, and r is the correlation coefficient between actual and predicted. The mean square error of prediction is decomposed into three components, $(\overline{A}c-\overline{P})^2$, $(S_{Ac}-S_P)^2$, and $2(1-r)S_{Ac}S_P$. The first component measures the extent to which the mean square error is due to squared bias, the second is the error due to the squared difference between the actual and predicted standard deviations, and the third is the error due to the lack of correlation between the actual and predicted observations (weighted by their standard deviations). For convenience, the table also reports the root mean square error, as well as the bias $(\overline{A}c-\overline{P})$.

Using the criterion of root mean square error of prediction and the equations including the war years, the F-M entry M_2 finished sixth in a field of twelve entries, not counting M_1 which was not in the

TABLE 1—ACCURACY OF CONSUMPTION PREDICTIONS, 1959–70 FROM REGRESSIONS ESTIMATED 1929–58
(billions of current dollars)

Independent Variable(s)	Root Mean Square Error	Mean Square Error	$(\overline{A}c-\overline{P})^2$	$(S_{Ac}-S_P)^2$	$2(1-r)S_{Ac}S_P$	$\overline{A}c-\overline{P}$
A. Estimation 1929–58						
M_1	161.8	26171.4	22460.8	3676.5	34.2	149.9
M_2	47.4	2250.8	2156.2	17.8	76.8	46.4
A	198.1	39256.6	35070.0	2751.0	1435.7	187.3
XA	107.2	11496.2	3710.7	6826.1	959.3	−60.9
ZA	46.5	2160.7	2100.3	0.6	59.8	45.8
XA&ZA	59.1	3487.3	976.5	2266.9	243.9	−21.2
A^*	28.8	826.9	662.4	116.4	48.1	29.7
A^{**}	121.2	14695.8	13549.9	1040.9	105.0	116.4
A^{***}	74.3	5520.3	5095.6	367.7	57.1	71.4
L	60.0	3594.3	3338.4	107.3	148.6	57.8
L'	19.8	390.7	308.7	21.0	61.0	17.6
L''	28.9	832.5	734.7	5.8	92.0	27.1
L'''	22.5	505.9	447.3	20.9	37.6	21.2
B. Estimation 1929–58 excluding 1942–46						
M_1	150.6	22681.2	19100.8	3545.2	35.2	138.2
M_2	37.9	1435.5	1350.4	7.1	78.1	36.7
A	134.0	17965.4	14792.0	1169.3	2004.2	121.6
XA	126.1	15893.6	6582.9	8306.3	1004.5	−81.1
ZA	20.2	407.6	304.0	39.5	64.1	17.4
XA&ZA	28.7	825.9	117.7	638.3	69.8	−10.9
A^*	18.2	330.3	2.8	276.8	50.7	1.7
A^{**}	76.6	5874.3	5287.6	464.7	122.0	72.7
A^{***}	43.8	1920.6	1704.6	154.1	61.9	41.3
L	28.6	816.8	648.1	7.2	161.6	25.5
L'	11.3	127.0	0.4	63.6	63.0	0.6
L''	11.7	136.8	36.7	3.9	96.2	6.1
L'''	11.0	120.0	4.3	76.6	39.2	2.0

original field. Since the A-M entries are a multiple regression on XA and ZA, and a simple regression on A^* (which is the sum of XA and ZA), the field might logically be defined to exclude the two regressions using XA and ZA separately. Also, the autonomous entry A was defined by F-M and universally branded by their critics as a poor bet even before the race was run. Excluding M_1, XA, ZA, and A from the race, M_2 finished fifth out of nine. The winner of the derby is L', one of the four autonomous variables defined by Hester.

If the performance of equations estimated excluding the war years 1942–46 is examined, the same general conclusions hold. Excluding the war years improves the performance of every equation except the one using XA. Excluding M_1, XA, ZA, and A, and using estimates based on the 1929–58 period excluding the war years, M_2 runs seventh out of nine, and L''' is now the winner.

The most obvious difficulty with the performance of these equations is their almost universal underprediction as can be seen from the last column of Table 1 which gives the bias. Several possible explanations for this bias come to mind.

Consistent with the monetarist view would seem to be an explanation going somewhat as follows. The regression of consumption on M_2 was meant to account for long-run effects. Since no effort was made to account for the dynamics of adjustment, it is not surprising that there are long strings of residuals of the same sign. If the war years are included in the period of estimation, the residual in 1958 was about \$29 billion. Beyond the period of estimation, the residuals rise to \$51 billion in 1960, and then decline to \$35 billion in 1967. In addition, it might be argued that the large residual of \$71 billion in 1970 is a consequence of the interest rate ceilings on both CD's and passbook time accounts, and that these ceilings were at least partly responsible for the residuals 1965–69 as well. Since interest ceilings were not really binding during the 1929–58 estimation period but were partially effective during much of the 1959–70 period, the M_2 of 1959–70 is simply not the same variable as the M_2 of 1929–58.

Another possible argument is that the war years distort the M_2 results because the equation was not estimated over a long enough time period. Excluding the years 1942–46 improves the performance of M_2 in the 1959–70 period both by increasing the constant term in the equation from -0.954 to 2.294, and by increasing the M_2 coefficient slightly from 1.304 to 1.325. If this process is carried even further by excluding 1942–49 from the estimation period, the constant drops slightly to 1.905 but the M_2 coefficient rises to 1.367, further improving the predictions 1959–70. These predictions, not reported in this paper, still have a substantial bias but the root mean square error is lower. These coefficient estimates may be compared to the F-M estimates (reported in their Commission on Money and Credit paper) of -1.198 and 1.351 for the constant and M_2 coefficient, respectively, using observations 1929–58, and 7.812 and 1.315, respectively, using observations 1897–1958. Using the longer time span raises the constant, but clearly not by enough to eliminate the bias in predictions 1959–70. From this examination of the coefficients it would appear that the monetarist explanation of the bias would have to rely primarily on the interest ceilings argument since acceptance of the long lags argument apparently has the unsatisfactory implication of acceptance of a model with substantial prediction errors in the same direction for at least a decade.[6]

[6] Another possible explanation is that the income elasticity of the demand for money is slightly below unity. In equations of the type estimated by F-M, the income elasticity differs from unity only by virtue of the

Many economists would no doubt argue that the underprediction using M_2 (or M_1) alone is the result of the failure to take account of the interest elasticity of the demand for money. Interest rates did, of course, rise substantially during the post-sample period. This explanation requires that the single equation approach be abandoned since at least one additional equation is required to explain the interest rate.

The Keynesian explanation of the bias would probably go as follows. First, H, A-M, and D-M all argued in their original papers that the war years should be excluded from the estimation period, and so the significant improvement in the predictions from excluding the war years should not surprise these authors. But even with the war years excluded from the estimation period, the Keynesian equations generally underpredicted 1959–70.

The explanation that immediately comes to mind for the underpredictions of the Keynesian equations is their failure to include taxes. It may be recalled that the F-M definition of autonomous expenditures, A, included the federal budget deficit whereas H, A-M, and D-M all made a special point of rejecting this aspect of the F-M definition on the grounds that tax revenues were clearly induced rather than autonomous. While H, A-M, and D-M might well have included a variable such as full employment revenues in the definitions of their autonomous variables, none of them in fact did so, and so none of them took any account of income tax rates or of the revenues produced by income taxes.[7] The underprediction of the Keynesian equations in the 1960's could be the result of the Kennedy and Johnson Administrations' tax cuts in the form of the reductions in personal and corporate income tax rates, liberalized depreciation, and the investment tax credit.

B. *Multiple Regression Results Using Money and Autonomous Expenditures*

Table 2 gives, except for the first line, the multiple regression coefficients for M_2 with each of the various definitions of autonomous expenditures excluding XA and ZA. For ease of comparison, the first line of the table gives the regression statistics for the simple regression of consumption on M_2. Comparing the multiple regressions with the simple regression, it can be seen that in every case except two the addition of an autonomous expenditures variable lowers the coefficient for M_2. The two exceptions are for A and A^{**} when the war years are included in the estimation period.

These two exceptions are the only cases in which the autonomous spending variable has a negative regression coefficient. For the variable A, this finding merely replicates the F-M finding of a negative coefficient for the 1929–58 period, a finding later confirmed by Meiselman, p. 17, for the 1961–67 period. But the regression coefficients for the other autonomous expenditures variables are not negative.

As can be seen from the last column of Table 2, the multiple regressions produce better predictions 1959–70 than the simple regression using M_2 alone for every definition of autonomous expenditures except for A, A^{**}, and A^{***} for the estimation period excluding the war years. For the multiple regressions, it is again true that the major problem is underprediction.

constant term, which in the estimates is in any event very close to zero, and so the income elasticity approaches unity as the constant term becomes of smaller relative importance with the passage of time. A logarithmic formulation, on the other hand, has a constant elasticity which might well be estimated to be less than unity. Although not directly relevant to the M_2 definition of money, it is suggestive that there is some evidence that the income elasticity of demand for M_1 is about 0.85 (see Poole (1970)).

[7] The A-J paper (discussed below) did include tests involving full employment expenditures and revenues.

TABLE 2—MULTIPLE REGRESSIONS 1929–58 AND THEIR PREDICTIVE POWER, 1959–70
(The dependent variable is consumption expenditures)
(billions of current dollars)

Autonomous Variable	Constant	M_2 Coefficient	Autonomous Variable Coefficient	R^2	Standard Error	Root Mean Square Error 1959–70
A. Estimation 1929–58						
—	−1.0	1.3	—	.946	18.8	47.4
A	−4.0	1.5	−1.0	.954	17.2	33.3
A^*	4.1	1.1	0.3	.946	18.7	40.6
A^{**}	−3.2	1.4	−0.2	.945	18.9	44.7
A^{***}	1.2	1.2	0.2	.945	18.9	47.3
L	0.0	1.3	0.1	.944	19.1	46.8
L'	5.7	0.9	0.4	.954	17.4	32.7
L''	4.2	1.0	0.4	.950	18.1	37.9
L'''	5.2	0.9	0.4	.953	17.5	34.0
B. Estimation 1929–58 excluding 1942–46						
—	2.3	1.3	—	.976	13.5	37.9
A	2.9	1.3	0.1	.975	13.8	39.1
A^*	31.7	0.3	1.2	.994	6.8	16.4
A^{**}	21.3	0.7	1.0	.984	11.2	53.7
A^{***}	22.0	0.5	1.1	.992	7.8	38.6
L	20.5	0.6	1.0	.987	9.8	30.0
L'	18.3	0.5	0.8	.994	6.5	15.0
L''	19.1	0.5	0.9	.994	7.0	17.8
L'''	19.1	0.5	0.9	.996	5.6	14.6

The dyed-in-the-wool Keynesian might interpret the multiple regression results as indicating that the inclusion of money in the analysis merely worsens the predictions since the root mean square error of prediction does not always decline when M_2 is added to the regression. We would not, however, accept this conclusion because, first, the addition of M_2 does reduce prediction errors for most definitions of autonomous expenditures and, second, even within the single equation context the multiple regression model appears to be misspecified. The difficulty is that in the formulation used by F-M the addition of an autonomous expenditures variable to the regression of consumption on money pushes the regression coefficient on money down to an improbably low value. An M_2 coefficient of 0.5, for example, implies that if the autonomous variable is held constant the ratio of consumption to money will gradually decline as the stock of money grows approaching the figure of only 0.5. That the equation is misspecified is supported by the fact that in every case the problem is one of systematic underprediction. A larger coefficient for M_2 would reduce or eliminate this problem.

C. *The Andersen-Jordan Model*

The original St. Louis Federal Reserve model was estimated by A-J on quarterly data 1952:1 through 1968:2. After reestimating this equation over the same sample period, and obtaining essentially the same coefficient estimates, we have examined the model's predictions over the thirteen quarter period 1968:3 through 1971:3. This model predicts the change in current dollar *GNP* from a distributed lag on the change in M_1 and a distributed lag on the change in high-employment government expenditures.

The entries for the postsample performance of this model corresponding to Table 1 are: root mean square error $=8.9$; mean square error $=78.4$; $(\overline{A}c-\overline{P})^2=16.8$; $(S_{Ac}-S_P)^2=1.7$; $2(1-r)S_{Ac}S_P=59.5$; $\overline{A}c-\overline{P}=-4.1$. These statistics, it should be remembered, are for the quarterly *change* in *GNP* in billions of current dollars. The root mean square error of 8.9 may be compared to the equation's standard error of estimate of 4.0.

Again there is a significant problem of bias, except that in contrast to the F-M equation the problem with the A-J equation is overprediction. Accumulating the first difference predictions to obtain level predictions, the overprediction amounts to 10.9 in 1968:4, 22.1 in 1969:4, 46.0 in 1970:4, and 53.9 in 1971:3. The A-J equation appears to be no more accurate than the F-M equation using M_2.

II. Significance of Findings

The policy implications of the results in Section I are clear. From the definitions of the autonomous expenditures variables it appears that none of the variables, including even (especially?) their government expenditure components, can be predicted very accurately a year in advance. This being the case, given a forecast of the nongovernment components of autonomous expenditures, the determination of government expenditures such that a desired level of consumption will be achieved will surely lead to substantially larger errors of prediction than indicated in the mean square error column of Table 1. On the other hand, even within the present institutional framework there can be little doubt that it is technically possible to set either M_1 or M_2 very accurately on a yearly average basis. Thus, the mean square error figure for M_2 in Table 1 is an appropriate measure of the probable accuracy of stabilization efforts using M_2 and based on the F-M formulation of the money-consumption relationship.[8] None of the relationships examined in this paper are accurate enough to be of much use in formulating discretionary policy. Nor are the relationships accurate enough to suggest that much confidence could be placed in a nondiscretionary policy rule based on one of these equations.

Table 1 lends considerable support to the argument that the war years should be excluded from the estimation period. For every independent variable tested, except for XA, the exclusion of the war years leads to an equation that predicts better during the 1959–70 period. From other regressions we have run (not reported here) there is also some evidence that the exclusion of 1947–49 as well as 1942–46 from the estimation period further improves the predictions beyond the period of estimation. Given the extent of the dislocations caused by the war, it would not be surprising if the postwar adjustments lasted several years and thereby distort estimates from regressions using those years.

Appraising these results in terms of testing rival theories, it would appear that the decision must still be rated a draw. By defining autonomous variables that produced $\bar{R}^2$'s as high as M_2 did in the sample period, H, A-M, and D-M blunted the F-M claim that . . . "The results are strikingly one-sided" (1963, p. 166). Given that during the postsample prediction period M_2 ran about in the middle of the pack, and given the disappointing performance of the A-J equation, the decision must still be rated a draw. Although some of the autonomous definitions performed better

[8] This statement, of course, assumes away possible problems of bias in the estimated equation, problems that might show up only during a period when M_2 was deliberately controlled by policy makers. During the 1959–70 period, M_2 was not a policy-determined variable. (This is not to say that M_2 was unaffected by policy but only that it was not systematically controlled by policy makers.)

than M_2, others performed worse. The fact that H, A-M, and D-M produced so many different definitions of autonomous expenditures, and that these definitions perform so differently in the projection period, clearly reinforces the F-M statement that . . . "However useful 'autonomous expenditures' may be as a theoretical construct, it is still far from having any generally accepted empirical counterpart" (1965, p. 754).

The findings support the contention that neither the simple Keynesian nor the simple quantity theory models provide an adequate understanding of business cycle fluctuations. Even supposing that the F-M equation comes back on track in the future, the equation could hardly be considered satisfactory considering the size and persistence of its errors over the 1959–70 period. The situation with respect to the autonomous expenditures equations is not really any better. With eleven different definitions of autonomous expenditures tested it is hardly surprising that several did quite well over a postsample period of twelve observations. There simply is not much empirical content in the single equation approach as employed in the studies examined here.

This discussion assumes that the appropriate method of testing a theory is to examine the accuracy of its predictions of endogenous variables given knowledge of the values of the variables specified by the theory to be exogenous.[9] Such a theory, it may be noted, may not be of much value in *ex ante* forecasting if the values of the exogenous variables cannot be predicted *ex ante*. Conversely, since it may be possible to make accurate short-run forecasts of an endogenous variable solely on the basis of knowledge of the cyclical and other properties of the time-series of the variable, short-run forecasting models may not be of much value in testing rival theories. Since the controversy being examined concerns rival theories, *ex ante* forecasting is not directly relevant.

During this controversy there has been little discussion of the appropriate testing methodology. In their CMC paper, F-M take the view that . . . "the question at issue is mainly the short-term stability of the relations being compared" (p. 174). Although this criterion seems to be in conflict with the F-M preference for using a long historical series, if short-run stability is the appropriate criterion, the 1959–70 performance of M_2 is quite good. Over the postsample period, the major problem with the M_2 equation is bias; as can be seen from the next to last column of Table 1, relatively little of the mean square error for M_2 is due to a poor correlation of M_2 with consumption.[10]

It can be argued, however, that the short-term stability of relations is not the appropriate criterion since with *U.S.* economic data almost any variable will be closely related to consumption over a given period if only the period is short enough. The levels of most, and the first differences of many, series exhibit substantial serial correlations around linear time trends, but the trends themselves differ substantially from one short period to the next. Two series correlated with each other will, therefore, have high short-run correlations. But of what value is a theory that asserts that over the next ten years the relationship between X and Y will be "close" and "stable" but that there is no way of knowing whether a 10 percent change in X will "cause" a 20 percent change in Y or a 5 percent change in Y? An adequate theory to explain Y must either involve a stable long-run relationship between X

[9] This statement should not be interpreted as claiming that accuracy of predictions is all that is relevant, but rather that no theory is acceptable if it is contravened by the facts.

[10] In the 1959–70 period the correlation between C and M_2 is .997.

and Y, or offer an explanation of why and how the relationship between X and Y shifts from one relatively short period to the next by adding additional explanatory variables to the theory. The argument that a relationship has broken down because of "structural shifts" is inherently nontestable unless the shifts themselves are specified so that their hypothesized effects can be tested on another body of data.

This argument does not deny that the period by period correlations between money and consumption (or income) are impressively high. The evidence is strong that money plays an important role in business cycle fluctuations. Yet the variability of the regression coefficient for money period by period makes clear the necessity of adding additional variables to obtain a satisfactory explanation of business cycle fluctuations.

Our findings emphasize the futility of the R^2 game. The F-M critics all came up with definitions of autonomous expenditures that produced regression R^2s rivaling or exceeding the R^2 for M_2. But beyond the period of estimation many of these relationships fall apart. Whatever one's biases in terms of the Keynesian and quantity theory approaches, these findings indicate that none of the single equation models predict the future very well. The F-M and A-J equations have large post-estimation biases in opposite directions while the H, A-M, and D-M equations present a very mixed picture. In terms of verified empirical knowledge, economists have good reason to be modest.

The key to the accumulation of empirical knowledge, as opposed to the construction of empirical hypotheses, is the replication of empirical studies on new bodies of data. Such work ought to be carried out on a routine basis in economics just as experiments in the natural sciences are replicated as a matter of course. But if there is to be systematic *ex post* review of empirical work, the original investigators must put their studies in such a form that they can be replicated and used to make predictions outside the sample period used for estimation. The F-M, H, A-M, D-M, and A-J studies are ideal in this respect. But a similar study of the simultaneous equations models of the *U.S.* economy is extremely difficult because these models are in a constant state of flux. It would be interesting, and probably sobering, to examine the 1970 predictions of a simultaneous equations model of 1960 vintage.[11]

If the F-M argument is interpreted not as a brief for a single equation approach, but as a brief for an empirically testable approach, then it would seem that the only significant fall-out so far from this controversy has been the St. Louis Federal Reserve model. The construction of large models has been extremely valuable, but it is unfortunate, in our view, that there has not been much more effort spent on small models that are subject to *ex post* testing and evaluation.[12] What this controversy has confirmed is that it is easy to meet the challenge of obtaining a high R^2 during the period of estimation, but not so easy to meet the challenge of post-estimation predictive accuracy. Our challenge is for someone to find a multi-equation model constructed a few years ago that has greater postsample predictive accuracy than the equations examined in this study.

[11] Even the within-sample performance of simultaneous equations models is not always very satisfactory. For example, the 1969 version of the FRB-MIT-PENN model went far off track in dynamic simulation over the 1956–68 period. See Poole (1971) pp. 164–66.

[12] The major complaint about large models has been that they are so complicated that it is impossible to fully understand their workings. This view has some validity but should not be overemphasized because with sufficient labor it is possible to understand these models. Much more serious, in our view, is the problem discussed in the text—that is, that these large models are not really subject to *ex post* evaluation, at least not without a tremendous expenditure of resources.

REFERENCES

L. C. Anderson and J. L. Jordan, "Monetary and Fiscal Actions: A Test of Their Relative Importance in Economic Stabilization," *Fed. Reserve Bank St. Louis Rev.*, Nov. 1968, *50*, 11–24.

A. Ando and F. Modigliani (1965a), "The Relative Stability of Monetary Velocity and the Investment Multiplier," *Amer. Econ. Rev.*, Sept. 1965, *55*, 693–728.

——— and ——— (1965b), "Rejoinder," *Amer. Econ. Rev.*, Sept. 1965, *55*, 786–90.

M. DePrano and T. Mayer (1965a), "Tests of the Relative Importance of Autonomous Expenditures and Money," *Amer. Econ. Rev.*, Sept. 1965, *55*, 720–52.

——— and ——— (1965b), "Rejoinder," *Amer. Econ. Rev.*, Sept. 1965, *55*, 791–92.

M. Friedman and D. Meiselman, "The Relative Stability of Monetary Velocity and the Investment Multiplier in the United States, 1897-1958," in *Stabilization Policies: Commission on Money and Credit*, Englewood Cliffs 1963.

——— and ———, "Reply to Ando and Modigliani and to Deprano and Mayer," *Amer. Econ. Rev.*, Sept. 1965, *55*, 753–85.

——— and ———, "Reply to Donald Hester," *Rev. Econ. Statist.*, Nov. 1964, *46*, 369–76.

M. Friedman and A. J. Schwartz, *A Monetary History of the United States*, Princeton 1963.

D. D. Hester (1964a), "Keynes and the Quantity Theory: A Comment on the Friedman-Meiselman CMC Paper," *Rev. Econ. Statist.*, Nov. 1964, *46*, 364–68.

——— (1964b), "Rejoinder," *Rev. Econ. Statist.*, Nov. 1964, *46*, 376–77.

D. Meiselman, "Stock of Money or Autonomous Expenditures: Some Recent Evidence," *Business Econ.*, summer 1968, *3*, 14–18.

K. E. Phillips, "The Short-Run Stability of Velocity and the Autonomous Spending Multiplier, 1946–1962," *J. Polit. Econ.*, May/June 1969, *77*, 418–29.

W. Poole, "Rules of Thumb for Guiding Monetary Policy," in *Open Market Policies and Operating Procedures—Staff Studies*, Washington 1971, 135–89.

———, "Whither Money Demand?" in A. M. Okun and G. L. Perry, eds., *Brookings Papers*, Washington 1970, *3*, 485–500.

Board of Governors of the Federal Reserve System, *Fed. Reserve Bull.*, Dec. 1970, *56*, 895–98.

U.S. Department of Commerce, *The National Income and Product Accounts of the United States, 1929–1965, Statistical Tables.*

———, *Surv. Curr. Bus.*, July 1968, July 1969, July 1970, July 1971, *48*, *49*, *50*, *51*.

On Revaluations versus Devaluations

By EDWARD S. HOWLE*

In reforming the system of pegged exchange rates, one matter of concern is the distribution of the burden of adjustment between revaluations and devaluations. The same amount of adjustment can be accomplished with various revaluation-devaluation combinations, but changing the combination has several effects. One important effect was noted in a report by the International Monetary Fund:

> Generally, and over a period of time, a predominance of exchange rate adjustments in one direction, taking account of the size of the adjustments and the relative importance of the currencies concerned, will tend to increase their number, since it will require such adjustments to be made by countries that are originally in a balanced payments position to a greater extent than if movements in both directions were reasonably matched.[1]

This point is explored in the analysis that follows. It is shown that, of all the various revaluation-devaluation combinations that could keep all countries in payments equilibrium, there is one particular combination that will minimize the number of exchange rate adjustments, with each adjustment weighted by the size of the induced trade changes that result. It is also demonstrated that, within a pegged-rate system, the particular revaluation-devaluation combination by which equilibrium is maintained affects not only the distribution of the burden of adjustment between countries, but also the total burden of adjustment. Thus, some systems would be less costly in the aggregate than others.

The paper begins by defining the adjustment pivot point; this concept is central in the analysis. There follows a discussion of the influence of the quantity of reserves upon the pivot point or upon the revaluation-devaluation combination. Then a formula is specified for the revaluation-devaluation combination (or bias) that will minimize the weighted amount of exchange rate change required over time by all pegged-rate countries. Finally, the analysis is applied to the 1960–71 period. A rough estimate is made of the reduction in trade adjustments that might result from the optimal revaluation-devaluation combination, as opposed to adjustment only through devaluations.

I. The Adjustment Pivot Point: An Analytical Tool

The pivot point can be explained by examining a world composed of countries A through K.[2] Each country has a constant rate of inflation determined by policies aimed at internal objectives. Thus, balance-of-payments considerations do not influence rates of inflation.[3] Assume that the maintenance of payments equilibrium in the long run requires that countries A through K have the average annual *relative* rates of exchange rate change (p_i values) in Table 1, stated in percentages.

* University of North Carolina at Chapel Hill. I am indebted to Clarence H. Adams, Jr., Dennis R. Appleyard, John E. Jelacic, James C. Ingram, Donald B. Keesing, Kenneth J. Kopecky, Carlos F. J. Moore, Hiroshi Ono, Steven S. Rosefielde, an anonymous referee, and others. Research support was provided by the Business Foundation of North Carolina, Inc.

[1] See *The Role of Exchange Rates in the Adjustment of International Payments*, p. 38.

[2] The pivot point concept was introduced independently by John Williamson, p. 3, and by Howle and Carlos F. J. Moore, pp. 430–31.

[3] These assumptions will be relaxed in Section II.

Assume further that a system of pegged exchange rates exists, and the band width about each parity is zero. Each country maintains payments equilibrium by periodic discrete devaluations or revaluations relative to gold and *SDR*s. The term parity will always mean the gold and *SDR* price of the currency, and a change in that price will be called a revaluation, devaluation, or parity change.

TABLE 1—RELATIVE ANNUAL AVERAGE PERCENTAGE RATES OF EXCHANGE RATE CHANGE (p_i VALUES) REQUIRED FOR LONG-RUN MAINTENANCE OF BALANCE-OF-PAYMENTS EQUILIBRIUM

p_i	Countries
0.0	*A*
2.0	*B, C, D*
4.0	*E, F, G*
6.0	*H, I*
8.0	*J*
10.0	*K*

In this example an infinite variety of revaluation-devaluation combinations could achieve the relative rates of exchange rate change in Table 1. To illustrate, adjustment could be accomplished by *A* maintaining a constant parity; *B*, *C*, and *D* each devaluing by 10 percent once every 5 years (an average annual rate of 2 percent); *E*, *F*, and *G* each devaluing by 10 percent once every 2-1/2 years; etc. In this case we would say that the adjustment pivot point is at zero on the scale of relative exchange rate changes, for the country at zero on that scale is the pivotal country; it does not adjust and all other countries adjust around it. Alternatively, the same relative rates of exchange rate change can be maintained by *A* revaluing by 10 percent once every 5 years; *B*, *C*, and *D* maintaining a constant parity; *E*, *F*, and *G* devaluing by 10 percent once every 5 years; etc.[4] In this case the pivot point is at 2 percent on the scale of relative exchange rate changes.

Thus, when all countries are maintaining payments equilibrium over time by some unique combination of average revaluation and devaluation rates, the adjustment pivot point, p^*, will have a numerical value equal to the average rate of *re*valuation for the country with the highest such rate.[5] Increasing the overall revaluation to devaluation ratio increases p^*. Also, as indicated in Table 1, p_i will mean the average percentage rate of exchange rate change of country *i* relative to the country with the highest rate of revaluation. A country's rate of parity change is $(p_i - p^*)$. For example, if p^* is equal to 6.0, country *B* must revalue at an average rate of 4 percent.

While p^* might be set at some given level by international agreement, the level of p^* might also be controlled by regulating the quantity of international reserves, provided that 1) no country with a p_i value of less than the desired p^* prefers holding infinite reserves to revaluing, and 2) a given percentage increase in reserves does not produce an equal or larger percentage increase in price.

To see how an increase in reserves might increase the value of p^*, we begin with a quantity of reserves so small that country *A* (Table 1), in preference to revaluation, will absorb all of the existing reserves that the other countries do not choose to hold. All of the other countries, since they have higher p_i values than *A*, will be forced to devalue periodically to avoid

[4] The error involved in adding or subtracting percentages can be made insignificant by the method used in Section III below, or by specifying the percentage change for a short time period, reducing the numerical size of the percentages.

[5] Exchange rate adjustment will involve a pivot point so long as adjustment is made vis-à-vis some *numeraire*. But, with freely fluctuating rates an no official reserves, there is no pivot point. Also, as Williamson has pointed out, p. 3, if all countries adjust around one key currency of a country that never initiates a parity adjustment, p^* must be equal to the p_i value for that country; a particular revaluation-devaluation combination is implied.

losing all of their reserves to A. Country A's balance of payments will be kept in equilibrium by the devaluations of other countries, and p^* will be equal to zero. To increase p^*, reserves must be increased enough so that A will prefer revaluation to holding all of the reserves that other countries do not choose to hold. So now let every country be allocated additional reserves. Deficit countries will be able to postpone devaluation, and it seems reasonable to assume that they will do so. In fact, those countries that must continue to periodically devalue might not choose to retain any of their share of the reserve increase. Most or all of the reserve increase will eventually tend to go to A. But if the increase is large enough, A will refuse to accept such a quantity of reserves, choosing instead to revalue each time its reserve quantity exceeds some standard. This is a shift in p^*. If B, C, and D are together willing to absorb all of the excess reserves that A refuses to absorb, then the new p^* value will be 2.0. In this way an increase in the quantity of reserves might increase the value of p^*.

Further analysis of the relationship between reserves and p^* will be presented elsewhere. For the present, it is sufficient to note the concept of p^* and the possibility of changing p^* either by international agreement or by less direct means such as changing the quantity of reserves.

II. The Exchange Rate Adjustment Model

It has been noted that a predominance of exchange rate changes in one direction tends to increase their number. In other words, an extreme p^* value, whether high or low, increases the amount of necessary exchange rate adjustment by all pegged-rate countries over a given time span. What may be less clear, but will be demonstrated below, is that when parity changes are randomly distributed through time, an extreme p^* value will increase not only the amount of needed exchange rate change but also the total amount of trade change that results from all adjustments. An extreme p^* value in the pegged-rate system causes the adjustment process to be accompanied by unnecessary trade balance shifts. And because of these effects of a high or a low p^* value, there must be some p^* value that will minimize the amount of exchange rate adjustment (however weighted), or the total amount of trade-balance change, required to keep all pegged-rate countries in equilibrium over time.[6] The exchange rate adjustment model presented here will explore this.

The assumptions used in the model can be slightly less stringent than those used in the earlier numerical example. In particular, we want to admit that efforts to offset the employment and inflationary effects of a parity change may be imperfect. It is therefore specified that p_i is country i's *long-run* average position on the scale of relative exchange rate changes, but there may be short-run deviations of the country's position from p_i. However, p_i, as a long-run value, is constant, and it is independent of the location of p^* or the quantity of reserves. The independence just mentioned need not imply that the overall world rate of inflation is unaffected by a change in p^* or the quantity of reserves; p_i values will be unaffected so long as relative rates of inflation remain unchanged. Thus we assume that only relative rates of inflation are unaffected by p^* and by the quantity of reserves in existence.

It is assumed, as before, that secular tendencies to balance-of-payments im-

[6] While the pivot point is a valid concept for any exchange rate adjustment system involving a *numeraire*, the analysis of the amount of trade adjustment accompanying the adjustment process is valid only for a pegged-rate system. When exchange rate changes are gradual, they may offset trade changes that would have otherwise occurred, rather than actually causing trade adjustments.

balances are corrected only by exchange rate parity adjustments, and between adjustments exchange rates are kept exactly at parities; the band width is zero. A country with p_i greater than p^* periodically devalues, while if the reverse is true periodic revaluation is necessary.[7] The average rate of parity change for country i is equal to the value of $(p_i - p^*)$. *Except where specified otherwise*, a country is assumed to base its decision to change its parity on its own payments and reserve situation, so that the decision is independent of concurrent parity action taken by other countries.

All values are denominated in units of a specific currency; a shift in p^* cannot permanently affect the value of one currency relative to another, but it does in time affect the value of every currency relative to reserves. For a given country, the size of the trade-balance change associated with a parity change is proportional to the size of the latter; elasticities are assumed to be the same for small and large parity changes by a given country. The adjustment of trade to exchange-rate changes is instantaneous.

We begin by using a numerical example to show that the value of p^* has no effect on the amount of induced trade change if parity changes are always coordinated in a particular way, but that p^* does otherwise affect the amount of induced trade change. Consider the world of six countries in Table 2.

TABLE 2—HYPOTHETICAL DISTRIBUTION OF p_i VALUES

p_i	Countries
0.0	*A, B*
2.0	*C, D, E, F*

In one way or another, *C* through *F* must depreciate their currencies relative to *A* and *B*, or *A* and *B* must appreciate relative to *C* through *F*, at an average annual rate of 2 percent if secular equilibrium is to be maintained. Assume initially that this is accomplished by *A* and *B* revaluing *simultaneously* by 10 percent once every 5 years. Let these revaluations result in a \$4 trade-balance deterioration for *A*, the same for *B*, and a \$2 improvement each for *C* through *F*. The total induced trade change, avoiding double counting, is \$8 per 5 years.

The amount of trade change resulting from parity changes can be this low only when parity changes are coordinated so that no country depreciates (appreciates) relative to another country that will later have to depreciate (appreciate) relative to the first. Let this be defined as perfect simultaneity. When adjustments are perfectly simultaneous, and only then, is the pivot point irrelevant to the amount of trade change taking place. If countries *C* through *F* had simultaneously devalued by 10 percent once every 5 years, rather than *A* and *B* revaluing, the relative change taking place is obviously unaltered, and the amount of induced trade change would still be \$8. But if adjustments are not perfectly simultaneous, the pivot point, p^*, affects the amount of trade-balance change required to maintain equilibrium, and regardless of the value of p^* the amount of such change will be greater than it will be with perfect simultaneity.

Consider, for example, the amount of trade adjustment when p^* is equal to 2.0 and adjustments are not coordinated. Countries *A* and *B* must each revalue at the same rate as before; let us assume that each revalues by 10 percent once every 5 years. With the adjustments uncoordinated, and with the assumption of instantaneous trade adjustment to parity changes, the probability is zero that the two adjustments will overlap. When the

[7] This follows from the definitions of p_i and p^*.

first revaluation takes place, it involves not only the necessary price changes between that country and the countries with p_i values of 2.0, but also an unnecessary price change between the acting country and the other country with the same p_i value. If the countries are of similar significance in trade, about 25 percent more trade is affected by the price change, and the total change in the trade balance might be as much as 25 percent larger than that attributable to the country's action in the simultaneous adjustment case. Of course, when the second of the two revaluing countries acts, it offsets the unnecessary price change between them, causing trade changes between them that reverse the initial ones. Both of these trade changes involve costs that are the result of nonsimultaneity.[8]

But, most important, observe the effect of changing p^* to zero when adjustments are not simultaneous. In that case four countries, rather than two, must adjust by 10 percent once every 5 years; with the countries of similar significance in trade about twice as much trade change will be induced as with nonsimultaneous adjustment and a 2.0 percent pivot. Each adjustment now involves a necessary depreciation relative to only two countries, and an unnecessary depreciation relative to the other three countries with identical p_i values. It is evident that, when adjustments are not perfectly simultaneous, changing p^* will alter the total amount of trade change accompanying the adjustment process.

We have seen that, with perfect simultaneity of the adjustment process, the value of p^* will affect the number of parity changes required for all countries to maintain equilibrium, but will not affect the resulting amount of adjustment-induced trade change. But without perfect simultaneity, there will be unnecessary trade changes, and these will be affected by the value of p^*. To generalize the latter result, we shall now derive the p^* value that will minimize the amount of unnecessary trade changes that would occur over time with nonsimultaneous pegged-rate adjustments. We begin by defining S_i as the absolute value of the trade change per percentage point parity change, resulting from i's independent (not simultaneous) parity adjustment; S_i can be stated as

$$(1) \qquad S_i = f(X_i, M_i, n_{si}, n_{di}, e_{si}, e_{di})$$

where X_i and M_i are the values of country i's exports and imports, n_{si} and n_{di} are the elasticities of supply and demand for i's exports, and e_{si} and e_{di} the corresponding import elasticities. For example, if A's revaluation of 10 percent produces a \$5 trade change, S_i must be equal to \$0.50.

The term $S_i|p_i-p^*|$ is the average trade change per time period (in country i and elsewhere) induced by country i's parity changes. This is because S_i is the absolute value of that trade change per percentage point of i's parity change, and

[8] It is not necessary for two p_i values to coincide for coordination of parity changes to reduce the cost of adjustment. If two periodic devaluers, for example, have different p_i values, unnecessary depreciations of one relative to the other will be avoided if a devaluation by the country with the lower average annual rate of devaluation is always accompanied by an equal or larger percentage devaluation by the other. As long as that requirement is fulfilled, unnecessary depreciations by either of the two countries *relative* to the other will not occur even though the country with the higher average rate of devaluation may have additional devaluations that are not coincident with devaluations by the other country. In general then, given a distribution of p_i values and a p^* value, the requirement for perfect simultaneity will be fulfilled if 1) whenever a country devalues, that devaluation is accompanied by a simultaneous devaluation of at least the same percentage by every country with a p_i value that is equal to or higher than that of the first country; and 2) whenever a country revalues, that revaluation is accompanied by a simultaneous revaluation of at least the same percentages by every country with an equal or lower p_i value. If these two conditions are fulfilled, there will be no unnecessary appreciations or depreciations of one country relative to another. A more detailed explanation and example is available from the author.

$|p_i - p^*|$ is the average amount of i's parity change per time period. The total value of all absolute trade changes per time period, that result from the parity changes required to keep all pegged-rate countries in payments equilibrium, is

$$T = \sum_{1}^{n} S_i \mid p_i - p^* \mid \tag{2}$$

Countries 1 through n are all pegged-rate countries.

Let R be the sum of S_i values for all countries with p_i values of less than p^* (all periodic revaluers), N be the sum of S_i values for those countries, if any, that have p_i values exactly equal to p^*, and D be the sum of S_i values for countries with p_i more than p^*. This means that

$$\sum_{1}^{n} S_i = R + N + D \tag{3}$$

From equation (2) it can be shown that the change in T resulting from an infinitely small change in p^* is

$$\frac{\Delta T}{\Delta p^*} = R + \frac{N\Delta p^*}{\mid \Delta p^* \mid} - D \tag{4}$$

where Δp^* is infinitely small. Equation (4) is not a derivative of equation (2) because the derivative of $S_i|p_i - p^*|$ is undefined where p_i and p^* are equal.[9]

It is found from equation (4) that T is minimized when[10]

$$\mid R - D \mid \leq N \tag{5}$$

This is based upon the assumption that S_i and p_i values are unaffected by changes in p^*. Of course, R, N, and D values are affected, for a change in p^* can shift a country from one category to another. Generally, one and only one p^* value will distribute countries among the R, N, and D categories in a way that will fulfill the condition stated in equation (5), and that p^* value will coincide with some p_i value so that N is positive. The only exception appears to be when there is a p^* value that makes R and D exactly equal when N is zero. In this case there will be one and only one range of p^* values, extending from a specific p_i to an adjacent one, that will fulfill the condition, and this range of p^* values will all produce the same (minimum) T value.

We now know the p^* value that will minimize the amount of unnecessary trade changes accompanying nonsimultaneous parity changes over time. Ignoring the problem of finite increments, it is the p^* value that will divide countries between revaluers and devaluers in such a way that the sum of S_i values for periodically revaluing countries (or R) is equal to the sum of S_i values for periodically devaluing countries (or D). This minimization condition can be understood intuitively by considering the effect of raising p^* by 1.0 percent. For simplicity assume that there is no p_i value such that

[9] Differentiation can be used except where p_i and p^* are equal. Since the derivative of a sum is equal to the sum of the derivatives, each $S_i|p_i-p^*|$ term in equation (2) can be differentiated separately to obtain

$$\text{(A)} \qquad \frac{\partial S_i \mid p_i - p^* \mid}{\partial p^*} = S_i \qquad \text{where} \quad p_i < p^*$$

$$\frac{\partial S_i \mid p_i - p^* \mid}{\partial p^*} = -S_i \quad \text{where} \quad p_i > p^*$$

Where p_i and p^* are initially equal, $\Delta S_i|p_i-p^*|/\Delta p^*$ is equal to S_i when Δp^* is positive, and to $(-S_i)$ when Δp^* is negative. This can be stated as

$$\text{(B)} \qquad \frac{\Delta S_i \mid p_i - p^* \mid}{\Delta p^*} = \frac{S_i \Delta p^*}{\mid \Delta p^* \mid} \quad \text{where} \quad p_i = p^*$$

The only case not yet considered is where (p_i-p^*) has a small nonzero value and the change in p^* causes the sign of (p_i-p^*) to change. But the change in p^* can always be made small enough to eliminate this possibility; this case can be ignored at the limit. Equations (A) and (B) can therefore be summed and the R, N, and D notation substituted to obtain equation (4).

[10] For a minimum to exist, neither an increase nor a decrease in p^* can decrease T. From equation (4) an increase in p^* will not decrease T so long as $(R+N-D \geq 0)$, and a decrease in p^* will not decrease T so long as $(R-N-D \leq 0)$. Thus the value of $(R-D)$ must lie between $(-N)$ and N, as is stated in equation (5).

($p^* \leq p_i < 1.01\ p^*$). Each of the countries with p_i less than p^* must increase their annual average rate of revaluation by 1.0 percent. The trade changes throughout the world associated with their revaluations rise by R per time period, for that is the total trade change per percentage point of their parity change. Similarly, the trade changes induced by parity changes of those above the pivot decline by D. Thus, the total adjustment-induced absolute trade change falls with a rise in p^* if R is less than D, is at a minimum if the two values are equal, and rises if R exceeds D. This confirms the equation (5) result.

Note what equation (5) means for the special case in which all countries have identical elasticities and average trade balances of zero over time. The ratio of S_i to (X_i+M_i) will be identical for every country, and the optimal pivot will be the one that causes countries to be divided between the revaluers and devaluers so that the total trade of periodic revaluers is equal to the total trade of periodic devaluers.[11] That will probably be approximately true even if these conditions are not met. *Thus, given the lack of accurate elasticity estimates, it would make sense for countries of the world to agree upon a system of exchange rate change in which countries below the pivot have roughly the same value of imports and exports as do those above.*

Finally, two points should be noted. It has been said that equation (5) states the condition for minimizing the amount of trade change accompanying the adjustment process. But, it is equally correct to say that equation (5) states the condition for minimizing the weighted number of parity changes, where the weighting is the amount of trade change that results from each parity change. The two criteria are definitionally identical. Also, note from equation (5) that it is the division of countries between those above the pivot and those below that matters, not how far above or below countries are. The optimal value of p^* will not be affected by a doubling of the rate of inflation, and a corresponding increase in p_i, of a country whose p_i value is already above the optimal p^* value.

This completes the theoretical analysis. We have shown that, if the adjustment process is perfectly simultaneous, the value of p^* will affect the number of parity changes required for all countries to maintain equilibrium, but it will not affect the total amount of trade change that results. Regardless of the pivot point there will be no unnecessary trade adjustments in this case. But without simultaneity, the adjustment process will involve unnecessary trade adjustments; a country will depreciate (appreciate) relative to other countries that will later find it necessary to depreciate (appreciate) relative to the first. There is a p^* value that will minimize (but not eliminate) the aggregate amount of unnecessary trade change that results from nonsimultaneity. That p^* value is the one that will fulfill the equation (5) condition, subject, of course, to the assumptions stated.

III. An Empirical Illustration

The analysis will now be applied to p_i values for the period December 31, 1960 through December 31, 1971. This will illustrate the use of the model and give a rough indication of how the amount of trade adjustment varies in relative magnitude with changes in p^*.

At the time of this writing exchange rate data for the end of 1971 was obtainable for 99 countries from an International Monetary Fund press release. We added Switzerland and the United States, giving

[11] Williamson, p. 14, suggests incorrectly that parity adjustment will be minimized if all countries adjust around the one with the modal rate of inflation.

TABLE 3—SELECTED TRADE-BALANCE CHANGE COMPUTATIONS[a]

			$S_i\|p^*-p_i\|$ values where p^* equals[b]				
(1) Country	(2) p_i	(3) S_i	(4) 0.00	(5) 18.36	(6) 26.62	(7) 30.20	(8) 69.27
(1) Germany	0.00	58.326	0	1,071	1,553	1,761	4,040
(2) Japan	10.95	29.064	318	215	456	559	1,695
(3) Australia (Gold)	18.36	9.559	176	0	79	113	487
(4) United States	26.62	86.478	2,302	714	0	309	3,688
(5) France	30.20	32.576	984	386	116	0	127
(6) United Kingdom	33.85	42.961	1,454	666	311	157	152
(7) Venezuela	54.02	6.614	357	236	181	185	101
(8) India	69.27	6.033	418	307	257	236	0
(9) 86 country total			111,312	55,805	53,368	61,278	223,971

[a] The p_i and p^* values are stated as percentages; all other values are in dollars divided by 10^7.
[b] The p^* values listed correspond to the following pivots: 0.00, Germany; 18.36, gold and *SDR*s; 26.62, United States; 30.20, France; 69.27, India.

us 101 countries that accounted for most of the trade outside of the Soviet area. Eight countries of this group, including Canada, were excluded because of flexible rates; the analysis does not apply to flexible-rate countries. Multiple-rate countries in general were included as long as there was one principal trade rate; only the Philippines and Viet-Nam were excluded because of multiple rates. Five very small countries were excluded because of nonexistent or doubtful data. This left a group of 86 countries in the study.[12] The 8 countries in Table 3 are representative and will illustrate the calculation methods used.

It will be recalled that p_i is the average percentage change in i's exchange rate per time period relative to the exchange rate of the country that revalues at the fastest average rate. Using the entire 11 years as the time period, we have computed each p_i value (Table 3, col. 2), as

$$p_i = \frac{\log (c_i/c_1)}{\log 1.01} \tag{6}$$

The term c_i is the number of units of i's currency per dollar at the end of 1971 over the number of units per dollar at the end of 1960.[13] And c_1 is the c_i value for Germany, the country with the greatest rate of revaluation; by definition p_1 is zero. Equation (6) produces the same result that would obtain if p_i values were first computed as ordinary percentage changes for a very short time period, then multiplied by the ratio of 11 years to that time period; the effect of compounding is removed. This has the advantage that it eliminates the error that would otherwise

[12] In order of p_i values: Germany, Netherlands, Japan, Switzerland, Austria, Belgium, Algeria, Australia, Iraq, Kuwait, Libyan Arab Republic, Malaysia, Morocco, Nigeria, Saudi Arabia, Singapore, Somalia, Ethiopia, Norway, Sweden, Italy, Portugal, Surinam, Netherlands Antilles, Hong Kong, Greece, Haiti, Honduras, Iran, Jordan, Kenya, Liberia, Mexico, Nicaragua, Pakistan, Paraguay, Panama, Sudan, Dominican Republic, El Salvador, Tanzania, Thailand, Uganda, Zambia, China, United States, Denmark, Ivory Coast, Malagasay Republic, Mauritania, Niger, Senegal, Togo, Upper Volta, Dahomey, France, Malta, South Africa, Syrian Arab Republic, Sierra Lione, Barbados, Cyprus, Gambia, Jamaica, Malawi, Spain, Ireland, United Kingdom, Burma, New Zealand, Tunesia, Guyana, Costa Rica, United Arab Republic, Finland, Venezuela, Nepal, Peru, India, Turkey, Ecuador, Ghana, Rwanda, Iceland, Israel, Zaire.

[13] For sources of data used in this section of the paper, see the International Monetary Fund, New York Times, and United Nations references.

be involved in comparing large and small percentages, and in adding or subtracting one percentage from another (i.e., p^* from p_i). Also, it prevents any bias that might result from using the entire 11 years as the time period. Calculating percentages by a simple mid-range base formula produced virtually identical results, however.

Next, S_i values (Table 3, col. 3) were computed for each country, the absolute change in the trade balance per percentage-point exchange rate change. This required export values, import values, and the usual elasticities. Mid-period (1966) export and import values were used, stated in dollars, in order to eliminate the problem of deflation. Export and import values were reduced to exclude intra-bloc trade for all countries changing rates along with sterling.

Reliable elasticity values, of course, are not available for the 86 countries. But our purpose is to obtain an index of relative amount of trade adjustment corresponding to different pivots, not an estimate of absolute amounts. This means that absolute elasticity values are not very important, but relative values can matter.[14] Moreover, it would do little harm to apply the same arbitrary elasticity values to all countries as long as elasticities were not correlated with p_i values. But no industrial countries are at the extreme upper end of the p_i scale, and no less developed countries are at the extreme bottom of the p_i scale. It does not seem likely that elasticities and p_i values are unrelated. To determine how much the possible relationship between elasticities and p_i values would affect our result, we made two sets of calculations. One used the same elasticities for all countries, and the other used the assumption that export and import values were twice as responsive to exchange rate changes by the 13 industrial countries in the study as by the remaining countries. Both results will be presented. The computation of S_i values in Table 3, column 3, uses the assumption that, for every country, the absolute value of the percentage change in exports per percentage-point parity change is 2.0, and the corresponding import value is 1.0. For example, Germany, with exports of \$20,145 million, and imports of \$18,036 million, had an S_i value of 0.02 times the former, plus 0.01 times the latter, or \$58.326x10[7].

Finally, columns 4 through 8 show how the adjustment-induced trade-balance change ($S_i|p^*-p_i|$) is computed for each country corresponding to various pivots. For example, the row 2, column 6 entry shows the trade-balance change that would be induced by Japanese revaluation if p^* corresponded to the *U.S.* p_i value of 26.62; that is, if Japan and all other countries adjusted to a constant *U.S.* parity. With p^* equal to 26.62 and Japan's p_i value equal to 10.95, the relative adjustment that actually took place would have required Japanese revaluation of 26.62 minus 10.95, or 15.67 percent. Multiplying this $|p^*-p_i|$ value by the Japanese S_i value of \$29.064x10[7] gives us an induced trade balance change of \$455x10[7].

The total trade adjustment for various p^* values is listed at the bottom of the table. With the elasticity assumptions used in this example, the adjustment minimizing p^* value is 26.62, which is the same as the *U.S.* p_i value. The amount of trade adjustment is lower for the 26.62 p^* value (col. 6) than for any other p^* value, not just those listed in the table. The relative exchange rate changes that took place, if they had occurred at random times within a pegged-rate system, would have involved the least trade adjustment

[14] Doubling the assumed responsiveness of trade to parity changes will double the calculated trade adjustment corresponding to each pivot, hence it will not alter the optimal p^* or the ratio of trade adjustments for one p^* relative to those for another.

TABLE 4—EFFECTS OF ALTERNATIVE ASSUMPTIONS

Assumptions	Optimal p^*	Relative amounts of adjustment if p^* is[a] 0	18.36	26.62	30.20	69.27
(1) Same as Table 1	26.62	2.086	1.046	1.000	1.148	4.197
(2) Cost based on exports only	26.62	2.062	1.038	1.000	1.157	4.228
(3) Cost based on imports only	26.62	2.131	1.061	1.000	1.132	4.136
(4) Less developed/developed elasticity weights 1:2	21.37	2.049	1.022	1.003	1.174	4.565

[a] The p^* percentages correspond to the following pivots: 0.00 Germany; 18.36, gold and *SDR*s; 26.62, United States; 30.20, France; 69.27, India. The figures given for relative amounts of adjustment are the amounts of adjustment for the given p^* values divided by the amount of adjustment for the optimal p^*.

if the United States had not adjusted and other countries had adjusted around the dollar pivot. The amount of trade adjustment would have been more than twice that if the same relative changes had ininvolved no revaluations (col. 4).

Table 4 summarizes the way in which the conclusions are affected by altering the assumptions. In rows 2 and 3 we have repeated the calculations, first on the assumption that only export revenues change, then only import revenues. Both results are nearly identical to those obtained earlier (row 1), showing that it is not important what ratio of export to import elasticities is used. In row 4 we assumed that trade was only half as responsive to parity changes in nonindustrial countries as in the 13 industrial countries in our study. Since the industrial countries tend to lie at the low end of the p_i scale, this change in assumptions lowers the optimal p^* value from 26.62 to 21.37. But the value is still above the 18.32 p_i value for gold and *SDR*s. Note that this change in the pivot is not of much importance, the use of the 26.62 pivot increases adjustment only three-tenths of 1 percent above its minimum.

In summary, the relative exchange rate changes that occurred during 1961 through 1971, if they had taken place independently within a pegged-rate system, would have caused about twice as much trade change with no revaluations as with the optimal pivot. But the precise location of the optimal pivot is not especially critical, either the gold-*SDR* pivot (18.32) or the dollar pivot (26.62) generally gives nearly optimal results.

IV. Summary and Implications

We have considered how the operation of the pegged-rate system is affected by the location of the pivot point. The pivot point affects the number of exchange rate changes, and, unless adjustment is perfectly simultaneous, it also affects the total amount of absolute trade-balance change, and the cost of such change, resulting from parity changes. The total amount of absolute trade-balance change was selected as the objective criteria to be minimized.

It was shown that, with the assumptions stated, the amount of induced trade change is minimized when the pivot is located so that the sum of S_i values for countries that tend to revalue is equal to the sum for countries that tend to devalue. If elasticities and average trade balances are unrelated to p_i values, this means that the optimal pivot is simply the one that makes the total trade of countries that tend to revalue equal to the total trade of countries that tend to devalue.

The empirical computations suggest that the total of absolute trade-balance

changes induced by parity adjustments might be half as great with the optimal pivot as with no revaluations. Even revaluation by a few countries at the low end of the p_i scale helps a great deal. One of the limitations of the numerical estimates, however, is that we have no way of knowing how changes in the pivot point would affect the formation of currency blocs that change their rates together.

It is perhaps of some historical interest that, with most assumptions tried, the dollar pivot would have resulted in minimum, or nearly minimum, trade adjustment during the 1961 through 1971 period. If other countries had been willing to adjust around the dollar pivot the amount of adjustment within a nonsimultaneous pegged-rate system would have been about minimum. (The adjustment that finally took place in 1971 was not within the pegged-rate system or at least had some degree of simultaneity, so the present analysis does not compare that amount of trade adjustment with the minimum pegged-rate trade adjustment.) Needless to say, countries with low p_i values cannot be expected to welcome efforts to raise the pivot point. But a few revaluations did occur before the 1971 disruption of the system, and indications are that even those few considerably reduced the amount of trade adjustment during the period, or, more likely, postponed the disruption of the system.

One last point. Some people may object to raising the pivot point on the assumption that a low pivot discourages inflation. The basis for this assumption is questionable; political pressure not to devalue may be stronger when devaluations are infrequent and unfamiliar, so that if Germany and the Netherlands had not revalued in the 1960's, devaluations might have become so commonplace as to reduce their political unpopularity. But in any case, it seems unlikely that the inflation argument would eclipse the argument for reducing trade adjustments. If the inflation argument is held to have merit, the costs of inflation and trade adjustment would have to be weighted against each other. The result would presumably be a pivot somewhat above the one that would minimize inflation, and somewhat below the one that would minimize trade adjustment.

REFERENCES

E. Howle and C. F. J. Moore, "Richard Cooper's Gliding Parities: A Proposed Modification," *J. Int. Econ.*, Nov. 1971, *1*, 429–36.

J. Williamson, *The Choice of a Pivot for Parities, Essays in Int. Finance,* no. 90, Nov. 1971, Princeton.

International Monetary Fund, *Direction of Trade*, No. 7, Annual 1966–70.

———, *Int. Financial Statist.*, various issues.

———, *Press Release*, no. 874, Dec. 30, 1971.

———, *The Role of Exchange Rates in the Adjustment of International Payments,* Washington 1970.

New York Times, Jan. 11, 1972.

United Nations, *Monthly Bull. Statist.*, various issues.

Human Capital: The Choice Between Investment and Income

By William J. Haley*

The process of investing in and accumulating human capital has been intensively studied in recent years. The fruit borne of this research can most clearly be seen in the way labor input is dealt with in nearly all empirical research now appearing in economic journals. Labor input is no longer measured as man-hours per unit of time, but an adjustment is made for the *quality* of labor input. In other words, the adjustment considers the human capital embodied in input, tending to make labor heterogeneous rather than homoegneous. The empirical benefits of dropping the assumption of homogeneous labor input is not without the substantial cost of making the theory more difficult.[1]

The literature in human capital has evolved along two essentially parallel patterns. The first, exemplified by Gary Becker, looks at individual investment and attempts to estimate the internal rate of return to that investment by equating properly discounted costs and returns. Although in theory the investments include a variety of activities which increase the individual's income stream in subsequent periods, most of the empirical work concerns estimating the rates of return to education. These studies have estimated rates of return to training individuals ranging from illiterates to candidates for Ph.D. and M.D. degrees; and one conclusion cannot be avoided—there is an economic explanation for the shape of life cycle earnings streams.

A second path along which the literature in human capital has evolved is exemplified by Yoram Ben-Porath (1967). The producer of human capital has a choice between producing additions to his stock or renting his stock in the labor market or both. He is assumed to make this choice in the process of maximizing his discounted future earning stream. In this literature, the empirical work has been rather meager, but not without reason. The principle observable phenomenon on which one must base verification of the theory is the observed income stream. The problem is then to deduce from the theory a theoretical income stream which can be compared to the observed. In general, this cannot be done. To produce an income stream from theory has been an intractable mathematical problem.

In what follows I will attempt to show the commonality between these two approaches to human capital, and in the process give one possible approach which avoids the mathematical problem referred to above. Two indirect tests are given of the model.

I. The Tie to Previous Work

One of the most important tenets of all human capital literature is that the

* Most of this work was carried out at North Carolina State University and was partially supported by National Science Foundation Grant No. GS2391. I would like to express my gratitude for helpful comments to T. Dudley Wallace, Mary Jean Bowman, Robert Fearn, Loren Ihnen, Thomas Johnson, and participants of the Econometrics and Human Resources Workshops at North Carolina State University. They, of course, are not responsible for any errors which remain. This is Michigan Agricultural Experiment Station journal article no. 5854.

[1] For example, if both capital and labor are heterogeneous the problem of aggregation, difficult to begin with, becomes even more difficult. See George Akerlof and Robert Solow.

cost of investing in oneself is measured by the earnings which are foregone during the investment period. The cost of a particular investment can be thought of as the difference between what could be earned if the investment did not take place and the actual earnings during the investment period. In particular, if we consider an individual who has made no previous investments in human capital, the cost of investment in period 1 (C_1) will be the earnings he could have made with his raw labor (X) less what he actually made in income during the period (Y_1). Thus, $C_1 = X - Y_1$. In the next period, if he continues to invest, the cost (C_2) will be what he could have earned with his raw labor (X) plus the return to the previous period's investment (r_1C_1) less his actual earnings (Y_2). Continuing with this line of reasoning, the cost of investment in period j is (see Becker, p. 65),

$$C_j = X + \sum_{i=1}^{j-1} r_i C_i - Y_j \tag{1}$$

Notice that if Y_j were combined with C_j on the left side, the sum $Y_j + C_j$ would represent the potential earning capacity of the individual. If Y_j were put on the left side and C_j on the right, equation (1) would represent earnings in period j. The sequence of values $\{Y_j\}$ for $j = 0, 1, \ldots, N$ would trace the earnings profile of a person of age N.

Becker and Barry Chiswick (1966) and Chiswick (1967) modify the model of equation (1) by defining cost in period j as a proportion of earning capacity. That is,

$$C_j = k_j Y_j^*$$

where

C_j = investment cost in period j
$Y_j^* = Y_j + C_i$ = potential earning capacity in period j
k_j = the proportion of earning capacity invested in period j

Making the substitution for C_j, equation (1) becomes,

$$Y_j^* = X + \sum_{i=1}^{j-1} r_i k_i Y_i^* \tag{2}$$

Now the choice variable is k_i. The X and r_i are exogenous to the individual, implying the time path of Y_j^* is determined by the choice of k_i made in each period. Once the path of Y_j^* is determined, the path of actual earnings can be deduced from the relation,

$$Y_j = (1 - k_j) Y_j^* \tag{3}$$

since actual earnings will be that percentage of earning capacity which is not invested. Equation (3) points out the fact that observed or disposable earnings is the proportion of earning capacity not being used for investment in period j. Clearly, the proportion of earning capacity devoted to investment is the choice variable from the individual's point of view and, therefore theoretical attention should be directed toward the determination of k_j over the life cycle.

Becker and Chiswick did not attempt to solve the theoretical question of how k_j is determined. With the help of an approximation[2] the closed form solution of equation (2) is found as

$$Y_j^* \simeq X e^{\sum_{j=1}^{j-1} r_i k_i} \tag{4}$$

[2] Chiswick shows that equation (2) can be written as

$$Y_j^* = X \prod_{i=1}^{j-1} [1 + k_i r_i]$$

Taking the logarithm of both sides,

$$\log Y_j^* = \log X + \sum_{i=1}^{j-1} \log [1 + k_i r_i]$$

Now noting that

$$\log [1 + k_i r_i] \simeq k_i r_i$$

is true for small $k_i r_i$, equation (4) follows.

where all terms are as previously defined. They assume the sequence $\{k_i\}$ to be positive during formal education and zero afterward. Figure 1 shows the shape of the sequence $\{k_i\}$ assumed by Becker and Chiswick. For any values of X and r_i, the time path of Y_j^* becomes constant after age S_3 and continues that way until the end of the earning life cycle. Therefore, their rather poor empirical results were not due to a bad model but rather the misspecification of the shape of the sequence $\{k_i\}$.

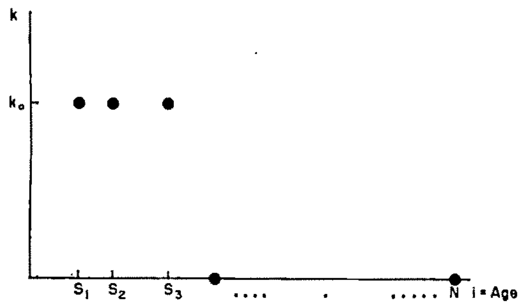


FIGURE 1. THE SEQUENCE $\{k_i\}$

Ben-Porath (1967) developed a model in which the determination of a continuous counterpart to the sequence $\{k_i\}$ could be studied. Although a closed form expression for this continuous counterpart, say $k(t)$, could not be derived, some of the characteristics of the function became apparent. Armed with this information, Thomas Johnson recognized the continuous formulation of equation (2) as a solvable integral equation whose solution was exactly equal to the continuous formulation of equation (4).[3] As with Becker and Chiswick, Johnson had a functional representation of earning capacity in terms of the proportion of earning capacity foregone as

[3] This simplified picture of Johnson's work is presented in order to show how the literature developed. Johnson solved his equation in completely general form with the rate of return, proportion of earnings foregone, and the rate of depreciation all having a time argument and an argument which included the personal characteristics of the investor.

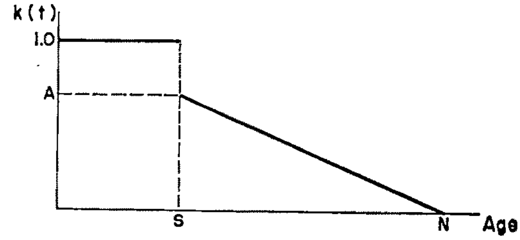


FIGURE 2. THE FUNCTION $k(t)$

well as the rate of return and rate of depreciation.[4] Using the results of Ben-Porath (1967) referred to above, Johnson assumed the function $k(t)$ had the relatively simple shape shown in Figure 2. Using this functional form for $k(t)$, Johnson's disposable earnings function had the requisite shape to articulate with observed age-earning profiles. Johnson's non-linear estimation procedure yielded estimates that seemed to make sense and his equation "fit" the data very well indeed. Certainly his procedure was a substantial improvement over Becker and Chiswick's method.

Notice that the function depicted in Figure 2 says that during formal education (ages less than S) the individual will forego his entire earning capacity in order to invest in human capital. Immediately following his formal schooling he will forego something less than 100 percent of his earning capacity for investment and, as time passes, this proportion will become smaller monotonically until, at the end of the earning cycle, his earning capacity and disposable earnings will be equal. These questions might be asked then: If an individual chooses between investment and earnings according to some objective within a choice theoretic framework, will $k(t)$ look like the one shown in Figure 2? Will it decline at a constant rate after age S? At age S, is the function $k(t)$ discontinuous—that is, does the individual evaluate human capital additions at a lesser

[4] The rate of depreciation has not been dealt with here for reasons of simplicity.

value after S than before as would seem to be implied by the discontinuity in $k(\mathrm{t})$ as shown in Figure 2? This, of course, is the neutrality hypothesis discussed in Ben-Porath (1970).

In what follows, a model which can give a theoretical answer to these questions will be formulated. In addition, a closed parametric form solution of the age-earnings profile implied by the model will be given. The economic implications of the model will be discussed as well as the results of empirical tests.

II. A Model of Optimal Human Capital Accumulation

The model proposed here will attempt to solve the problem of determining an individual's optimal stock of human capital at any given time in the life cycle, and how it should be maintained over time. It will be assumed that the problem will be solved by maximizing a particular objective function. From such a model, functions such as the path of human capital accumulation, the proportion of earning capacity foregone, and the earnings path over the life cycle can be derived in parametric form (provided the differential equations can be solved).[5]

Consider an individual or the typical individual of a group with an earning capacity of

$$Y^*(\mathrm{t}) = RE(\mathrm{t}) \tag{5}$$

where

$Y^*(\mathrm{t})$ = the individual's earning capacity at t

R = the constant rental rate per unit time on the stock of human capital

$E(\mathrm{t})$ = the total stock of human capital possessed by the individual age t

[5] For other efforts along these lines see Ben-Porath (1967), J. Ross Millar, H. Oniki, Sherwin Rosen, Eytan Sheshinski, and Yoram Weiss (1971b).

At any time t, he has the choice of renting all or any portion of his human capital stock in the labor market.

It will be assumed here that the only alternative use for his human capital is producing more human capital according to:

$$q[K(\mathrm{t})] = \beta K(\mathrm{t})^b \tag{6}$$

where

$q[K(\mathrm{t})]$ = the output of new human capital produced in period t

$K(\mathrm{t})$ = the portion of the existing stock of human capital used as input to produce more human capital in period t

β = the measure of the individual's ability to produce human capital, assumed constant over his life cycle

b = the scale parameter in the production function and is contained in the *open* unit interval

Assume, also, that capital markets are perfect. The individual's disposable earnings in any period will be the difference between his earning capacity and the cost of input used in the production of new human capital during that period. Disposable earnings, $Y(\mathrm{t})$, will be

$$Y(\mathrm{t}) = Y^*(\mathrm{t}) - RK(\mathrm{t}) \tag{7}$$

It will be assumed that the stock of human capital deteriorates at a constant and exogeneously given rate δ. The human capital stock will grow at a rate governed by

$$\dot{E}(\mathrm{t}) = q[K(\mathrm{t})] - \delta E(\mathrm{t}) \tag{8}$$

The choice facing this individual is quite clear. Since it is assumed he has no physical assets, his income-earning capacity is directly related to the size of his stock of human capital, which can grow only if he is willing to postpone part or all of his current earnings as the cost of growth. If it is assumed the individual has perfect

knowledge and there is no risk associated with his potential investment, then he will undertake the investment only when the discounted present value of the additional earning capacity attributable to the investment exceeds the value of earnings postponed during the investment period. Extending this logic, the individual's objective would be to maximize J the present value of his disposable earnings over the life cycle;

$$J = \int_0^N e^{-rt} Y(t)dt \tag{9}$$

where

N = the end of the earning life cycle
r = the rate at which the individual discounts future earnings
$Y(t)$ = disposable earnings or the difference between earning capacity and investment costs at time t

The evolution of disposable earnings over the life cycle is governed by the differential equation (8). The general form of the optimization problem is to maximize equation (9) subject to equation (8) and all the boundary equations. An explicit representation of the boundary conditions are

$$\begin{aligned} K(t) &\geq 0 \\ E(t) &> 0 \\ E(t) - K(t) &\geq 0 \end{aligned} \tag{10}$$

Application of Pontryagin's Maximum Principle after incorporating constraints (10) and the initial stock of human capital, E_0, assures the necessary conditions for a unique maximum to equation (9) with transversality conditions,

$$\lambda(N) \geq 0 \tag{11}$$

$$\lambda(N)E(N) = 0 \tag{12}$$

The interpretation of $\lambda(t)$ is that it represents the discounted marginal value of an additional unit of human capital at time t and, therefore, is interpreted as the individual's "demand price" for human capital.[6] It seems reasonable that at the end of one's earning lifetime, the marginal value of an additional unit of human capital for the purposes of generating income will be zero.[7] This means equation (11) becomes an equality.

Solving the differential equations which represent necessary conditions, the solution turns out to have two stages. The reasons for this will become clear as the discussion proceeds. In the technical jargon of control theory, the constraints force a boundary solution in the first stage whereas in the second stage the constraints do not affect the maximum thus allowing an interior solution. Both stages will be discussed beginning with the second stage.

In the second stage of the investment life cycle, the individual is producing less human capital than he is physically capable of producing (i.e., $K(t) < E(t)$). This means that the marginal cost of producing an additional unit of human capital exceeds the discounted value of additional income generated over the life cycle by that unit of human capital. Solving for $\lambda(t)$, this discounted marginal value of human capital in the second stage is,

$$\lambda(t) = \left(\frac{R}{r+\delta}\right)\left[e^{-rt} - e^{-(r+\delta)N+\delta t}\right] \tag{13}$$
$$t^* \leq t \leq N$$

where t* represents the beginning of stage two. The marginal value of human capital to the individual is independent of the quantity produced as was shown by Ben-

[6] See Kenneth Arrow. The term demand price was first used by Ben-Porath (1967).

[7] It can, no doubt, be effectively argued that certain activities which generate human capital in year N have some value. The values they have, however, satisfy utility through leisure or some other nonpecuniary argument in the utility function and not through the marginal increase in the income stream, provided year N is the last year in which the individual can earn income.

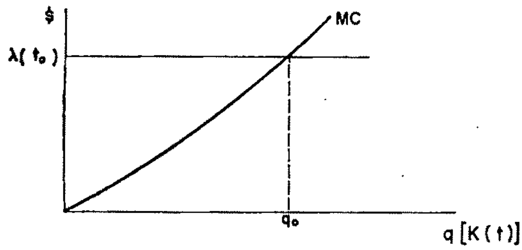


FIGURE 3. EQUILIBRIUM PRODUCTION OF HUMAN CAPITAL

Porath (1967). The quantity of human capital produced will be determined by the intersection of equation (13) with the marginal cost schedule shown in Figure 3. Over the life cycle, the marginal value of additional human capital decreases until age N, the end of the earning life cycle, when it is zero by assumption. Effectively, $\lambda(t)$ slides down the stationary marginal cost curve over the earning life of the individual.

The shape of the optimal path for $K(t)$ must now be determined given that production of human capital is determined by its marginal value. The optimal path of the human capital input to production over the second stage of this earning life cycle is:

$$(14)\quad K(t)=\left[\frac{b\beta}{r+\delta}\right]^{1/1-b}\left[1-e^{(r+\delta)(t-N)}\right]^{1/1-b}$$
$$t^*\leq t\leq N$$

where b and β are as previously defined in equation (6). A priori, it would be expected that if the marginal value of an additional unit of human capital is zero (see equation (12)), none would be produced. Equation (14) makes this expectation explicit since $K(N)=0$. The surprising fact about equation (14) is that after the individual stops specializing in the production of human capital (i.e., after stage one), the amount of capital used in the production of more human capital declines *strictly* monotonically.

The principal cause of the shape of $K(t)$ is the declining marginal value of an additional unit of human capital as shown by equation (13). Note, however, that production of human capital continues, albeit at a diminishing rate, until the end of the earning life cycle. If the life cycle were infinite, $K(t)$ becomes independent of t.

It is now possible to observe the optimal human capital accumulation path. In order to display this equation a convenient simplifying assumption of $b=1/2$ is made.[8] With this assumption the optimal path of human capital accumulation during stage two is

$$(15)\quad E(t) = Be^{-\delta t}+\frac{\beta^2}{2\delta(r+\delta)}\left[1-\frac{\delta}{r+2\delta}e^{(r+\delta)(t-N)}\right]$$
$$t^*\leq t\leq N$$

where

$$B=\left[\frac{\beta}{\delta}+\left(E_0^{1/2}-\frac{\beta}{\delta}\right)e^{-\delta t^*/2}\right]^2 e^{\delta t^*}-\frac{\beta^2}{2\delta(r+\delta)}\left[1-\frac{\delta}{r+2\delta}e^{(r+\delta)(t^*-N)}\right]e^{\delta t^*}$$

Multiplying equation (15) by the constant rate of return, R, the optimal path of earning capacity, $Y^*(t)$, is determined. Substituting equation (5) into equation (7), disposable earnings over the life cycle is

$$(16)\qquad Y(t) = RE(t) - RK(t)$$

which, after using equations (14) and (15), is determined in parametric form.

[8] The equation for optimal human capital accumulation can be solved for any real number b contained in the unit interval. For all such b when $b/(1-b)$ is an integer, the solution for $E(t)$ is a finite series as in the case given above for $b=1/2$. In the case where $b/(1-b)$ is not an integer, the solution for $E(t)$ is an infinite series. As long as b is contained in the unit interval the convexity of $E(t)$ remains unchanged as b varies. Therefore, using $b=1/2$ only makes the problem easier without making the results discussed below in any way less general. However, for certain kinds of empirical work, it is important that b be allowed to vary.

Equation (16) can be rewritten as

$$Y(t) = [1 - k(t)]RE(t) \tag{17}$$

where $k(t) = K(t)/E(t)$ is the proportion of earning capacity foregone to make current investments in human capital.

This is precisely the function which appeared in Becker and Chiswick, Johnson, and Ben-Porath (1967) as discussed in Section I. Notice that without a specific maximizing process such as equation (9), any function (or sequence in the discrete case) bounded above by one and below by zero over the life cycle, is admissible. Once the assumption is made that the individual invests in himself for the exclusive purpose of increasing his income, $k(t)$ becomes tied to zero at the end of the earning life cycle.[9] The behavioral assumption that he acts as if to maximize the discounted value of disposable earnings allows the determination of a unique function of $k(t)$. Notice that $k(t)$ is uniquely determined as a result of assuming the economic unit behaves according to well-recognized micro-economic principles. The actual functional form of $k(t)$ is obtained for stage two by the ratio of equation (14) to (15). Once this ratio is formed, it is easily discovered that $k(t)$ is a strictly monotonically decreasing function of time.

At this point, it becomes clear why the constraints force a multiple solution to the maximizing problem, one on the interval $(0, t^*)$ which is called stage one and a second on the interval (t^*, N) which is referred to as stage two. Using the functional form of $k(t)$ defined as above on stage two we find that $k(t)$ will exceed one for all values of t less than t^* or all values of t contained in stage one except t^*. But recalling that $k(t)$ is the proportion of the individual's human capital stock used in the production of more human capital, it is *physically* impossible for $k(t)$ to exceed one. An individual cannot use more than 100 percent of his stock. Apparently, on $(0, t^*)$, the value of a marginal addition to the stock exceeds the cost of making that addition, but the input supply is fixed in each period and will not allow the individual to increase production to the point where the marginal cost of adding more to the stock equals the marginal value of the addition. Figure 4 shows how these constraints operate. The long-run marginal cost (LMC) is the envelope of all short-run marginal cost curves (represented by MC_1, MC_2, MC_3) where it is assumed in Figure 4 that the parameter $b = 1/2$. The problem is that in the first two periods (represented by the subscripts 1, 2), the marginal value of additions to an individual's stock (MV_1 and MV_2) exceed the marginal cost of these additions, but due to the fixed supply of stock in these two periods he cannot produce at the output which would equate these two values. He will, of course, use all his available stock as input in the first two periods, but in the third period the availability of input is no longer a problem. The marginal value of additional stock has been declining with the passage of time (see equation (13)) and in the third period it becomes possible for the individual to equate the marginal cost of production

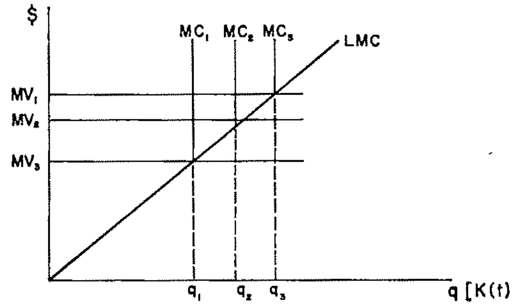


FIGURE 4. OUTPUT CONSTRAINTS WHEN MV OF AN ADDITION TO THE STOCK EXCEEDS MC

[9] At this point it may be worth reemphasizing that the individual has only two alternatives in this model, work and investment in human capital. Clearly leisure is a third alternative which for purposes of this paper is being ignored.

with marginal value of the addition to his stock. Thus, in the third period and all subsequent periods, his production will be constrained by cost considerations rather than an input fixity. The exact time that this switch takes place is represented by t*.

So stage one turns out to be an instance of specialization in production of human capital since the individual will follow a single-minded program of investing to the maximum of his capabilities. He will use his entire stock of human capital as input to the production function and forego all his earning capability during this time. This process will continue until he reaches the point at which the marginal cost of production exceeds the marginal value of that production and production is reduced. The reduced production will be observable as an increase in the amount of human capital the individual is willing to rent in the labor market.

It is now possible to define $K(\mathrm{t})$, $E(\mathrm{t})$, $k(\mathrm{t})$, and $Y(\mathrm{t})$ over the entire earning life cycle. This provides an opportunity to summarize all the important functions discussed above. All symbols used are as previously defined and $b=1/2$.

The path of human capital stock used to produce more human capital over the life cycle is[10]

$$(18)\qquad K(\mathrm{t}) = \begin{cases} E(\mathrm{t}), & 0 \leq \mathrm{t} \leq \mathrm{t}^* \\ \dfrac{\beta^2}{4(r+\delta)^2}\left[1 - e^{(r+\delta)(\mathrm{t}-N)}\right]^2, & \mathrm{t}^* \leq \mathrm{t} \leq N \end{cases}$$

Recalling that

$$B = \left[\frac{\beta}{\delta} + \left(E_0{}^{1/2} - \frac{\beta}{\delta}\right)e^{-\delta \mathrm{t}^*/2}\right]^2 e^{\delta \mathrm{t}^*} - \frac{\beta^2}{2\delta(r+\delta)}\left[1 - \frac{\delta}{r+2\delta}e^{(r+\delta)(\mathrm{t}^*-N)}\right]e^{\delta \mathrm{t}^*}$$

the optimal path for total human capital accumulation is given by equation (19). The proportion of human capital devoted to the production of more human capital is given by equation (20), and the optimal earnings path over the life cycle is given by (21).

[10] An appendix of these derivations will be provided upon request.

$$(19)\qquad E(\mathrm{t}) = \begin{cases} \left[\dfrac{\beta}{\delta} + \left(E_0{}^{1/2} - \dfrac{\beta}{\delta}\right)e^{-\delta \mathrm{t}/2}\right]^2, & 0 \leq \mathrm{t} \leq \mathrm{t}^* \\ Be^{-\delta \mathrm{t}} + \dfrac{\beta^2}{2\delta(r+\delta)}\left[1 - \dfrac{\delta}{r+2\delta}e^{(r+\delta)(\mathrm{t}-N)}\right], & \mathrm{t}^* \leq \mathrm{t} \leq N \end{cases}$$

$$(20)\qquad k(\mathrm{t}) = \begin{cases} 1.0, & 0 \leq \mathrm{t} \leq \mathrm{t}^* \\ \dfrac{\dfrac{\beta^2}{4(r+\delta)^2}\left[1 - e^{(r+\delta)(\mathrm{t}-N)}\right]^2}{Be^{-\delta \mathrm{t}} + \dfrac{\beta^2}{2\delta(r+\delta)}\left[1 - \dfrac{\delta}{r+2\delta}e^{(r+\delta)(\mathrm{t}-N)}\right]}, & \mathrm{t}^* \leq \mathrm{t} \leq N \end{cases}$$

$$(21)\qquad Y(\mathrm{t}) = \begin{cases} 0.0, & 0 \leq \mathrm{t} \leq \mathrm{t}^* \\ \dfrac{R\beta^2(2r+\delta)}{4\delta(r+\delta)^2} + RBe^{-\delta \mathrm{t}} + \dfrac{R\beta^2}{2(r+\delta)^2}\left(\dfrac{\delta}{(r+2\delta)} - \dfrac{1}{2}e^{(r+\delta)(\mathrm{t}-N)}\right)e^{(r+\delta)(\mathrm{t}-N)}, & \mathrm{t}^* \leq \mathrm{t} \leq N \end{cases}$$

The model is a function of essentially four parameters and an initial stock of human capital, which will be referred to as a parameter in all subsequent discussions. Since the rental rate, R, is essentially a factor of proportionality and has an unambiguous effect, the approach will be to discuss the other four parameters as they affect the capital accumulation path and the path of capital input. In particular, the effect of these four parameters on the length of time the individual spends specializing in the production of human capital and on the location of the time period in which the human capital accumulation path attains its maximum value will be discussed. Special attention will be given to $K(t)$ and $E(t)$ since they in turn determine the effect of parameter variations on $k(t)$ and $Y(t)$.

Figure 5 is a graphical picture of the shapes of $K(t)$, $E(t)$, and $k(t)$. Notice that the value of $E(t)$ and $K(t)$ is to be read off the left vertical axis and the value of $k(t)$ is to be read off the right vertical axis. The disposable earnings stream, $Y(t)$, will be discussed later.

One of the important characteristics of the model is the fact that t^*, the point when the individual stops specialization in the production of human capital, is determined endogenously as can be seen from Figure 5. Also, notice that the time when the stock of human capital is maximum occurs at t_m. At time t_m the rate at which the stock depreciates equals the rate at which it is being produced.[11] This is not a

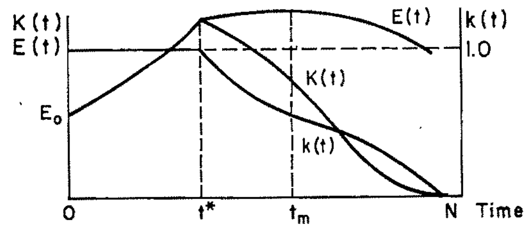


FIGURE 5. A GRAPH OF $E(t)$, $K(t)$, AND $k(t)$

particularly unusual result[12] and is explained by the fact that sometime after specialization in production stops the depreciated human capital exceeds the amount of newly produced human capital and $q[k(t)]$ becomes less than $\delta E(t)$ making $\dot{E}(t)$ in equation (8) negative. The point in time when $\dot{E}(t)$ goes from positive to negative is t_m.

Since the effects of β, the ability-to-produce parameter, and E_0, the initial endowment of human capital, on t^* are most interesting, let us discuss these first. By noticing that during stage two of the life cycle, (t^*, N), $K(t)$ is independent of E_0, it can be seen from Figure 5 that *ceteris paribus*, the larger one's initial endowment of human capital, the earlier he will stop specializing in the production of human capital and begin earning positive income. Further, the greater the ability index, β, the longer the period of specialization, other things constant. The size of one's initial endowment and his ability to produce tend to operate in different directions with respect to the length of time spent specializing in the production of human capital and the outcome is not clear a priori. This is contrary to previous theories.[13]

[11] To show that the maximum occurs at t_m, the first derivative of equation (19) on (t^*, N) obtains,

$$E'(t) = -\delta B e^{-\delta t} - \frac{\beta^2}{2(r+2\delta)} e^{(r+\delta)(t-N)}$$

The usual procedure is to set $E'(t)=0$ and solve for t. Since the second term on the right-hand side of $E'(t)$ is strictly negative for all values of t, in order for a maximum to exist in the interior of the interval (t^*, N), B must be negative. The sign of B is not at all easy to establish analytically, but can be established by using a computer to compute B for various parameter values. This was done, and it was found that B is indeed negative. Therefore, $E(t)$ has a maximum value in the interior of (t^*, N).

[12] See Becker (Appendix B) for a somewhat different view of how the stock of human capital is accumulated. However, he also concludes the stock path attains a maximum after full time investment has stopped.

[13] See Ben-Porath (1967, p. 358). Also, Weiss (1971a) criticized Mincer (1970) for assuming ability to be

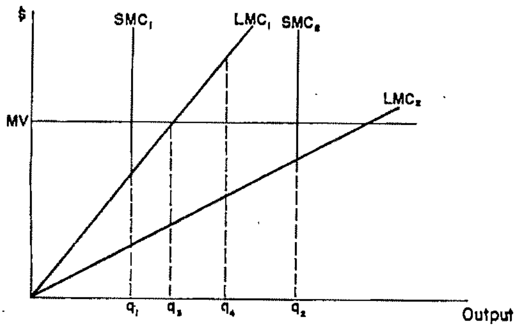


FIGURE 6. EFFECT OF E_0 AND β ON PRODUCTION

The phenomenon described above can be shown most clearly by considering two individuals who are identical except in their ability to produce human capital as shown by Figure 6. Since the higher the β the lower the marginal cost curve, individual 1, with LMC_1, will have a smaller ability index than individual 2. If the output using all the initial endowment as input for both individuals were q_2, then individual 1 would not specialize, producing at q_3, whereas individual 2 would specialize, producing at q_2. Clearly, the size of the ability index, β, is positively related to the time t* when specialization stops.

Now consider two individuals who are identical except for their initial endowment of human capital. In this case, let individual 1 have the smaller endowment. If he used it all for production, he could produce output q_1. Individual 2, on the other hand, could produce output q_4 if he used his entire initial endowment for production. If LMC_1 represents the long-run marginal costs for both individuals, 1 is faced with a short-run factor fixity and will specialize in production, whereas 2 will equate MV with LMC_1 at q_3, an output which is not constrained by an input fixity. Thus, individual 2 will not specialize and E_0 is negatively related to t*.[14]

The rate of discount will be inversely related to the time spent specializing. This result makes eminently good sense. The higher the discount rate, the less value an individual places on future dollars relative to present dollars. The individual with the relatively higher discount rate will be less inclined to forego present income for investment purposes than would the individual with a relatively low discount rate. Therefore, if all other parameters are the same, the individual with the higher discount rate will stop specializing in the production of human capital sooner than the individual with a lower discount rate.

Finally, an increase in the rate of depreciation will decrease the length of the period of specialization. The economic argument here is very analogous to that given above for the discount rate. The individual will have a larger present value of actual earnings if he places more importance on earnings early rather than later in the planning horizon because a high depreciation rate will eat away more of his stock of human capital late in the planning horizon than will a low depreciation rate. Early earnings will be more important in the objective function with a high rate of depreciation than with a low rate of depreciation, other things the same.[15]

The optimal disposable earnings path represented by equation (21) has a shape which is at first increasing, then begins to

positively related to t*. According to Weiss' model, ability can be either positive or negatively related to t*.

[14] These results are both independent of the assumption that $b=1/2$ as long as b is the same between individuals and is not equal to one or zero. If b is not equal between individuals the results may not hold.

[15] Analytically this proposition is not easily established. The difficulty is that $K(t)$ is negatively related to δ on both (0, t*) and (t*, N). In Figure 5 the reader can imagine both segments of $K(t)$ swinging downward. The location of t* depends on which segment has the larger relative shift for a given increase in the rate of depreciation. By computer simulation it was determined that the segment of $K(t)$ on (t*, N) shifted down relatively more than $K(t)$ on (0, t*) giving the result that an increased rate of depreciation decreased the interval of specialization, *ceteris paribus*. See my dissertation for these and other simulation results.

decrease after attaining a maximum. The maximum of the disposable earnings path is attained later in the life cycle than is the maximum of the human capital stock.[16]

The path of proportion of earning capacity foregone function $k(t)$, is displayed in Figure 5. A quick comparison of this function with the one used by Johnson represented by Figure 2 shows some substantial differences. In addition to the obvious differences in second order properties, the function generated by this model shows no notch at t^* (or S in Johnson's terminology). Johnson had no theoretical reason for a priori assuming $k(t)$ to be discontinuous at t^*. This was simply an empirical result which made his model fit the data better than a sectionally continuous $k(t)$ implied by this model. This empirical test may be an illusion created by the fact that using the end of formal schooling as t^* is incorrect. Possibly t^* occurs substantially before that time and using S rather than the true t^* causes the discontinuity.

If, however, S is a reasonable measure of t^*, this may be evidence of what Ben-Porath (1970) described as nonneutral human capital accumulation. Simply stated, it means that human capital produced after t^* has less value, after adjustment for the fact that it is produced later in the life cycle, than that produced before t^*. The neutrality hypothesis poses a serious question and it would appear that this is at least one way of considering it.

[16] One case in which the two paths, $Y(t)$ and $E(t)$, obtain maxima at the same time is when the scale parameter b is one. When $b=1$ the marginal cost schedule is constant and the individual will specialize as long as the marginal value of additional stock exceeds the marginal cost of production, but when the reverse is true, *all* production stops and the individual rents his entire stock in the labor market. This would imply a $k(t)$ that is constant at one until t^* and is constant at zero for all t after t^*. Clearly both $Y(t)$ and $E(t)$ would attain maxima at t^*. Ben-Porath (1967) also observed this limiting behavior.

III. Predicted Training Costs

The choice variable in this model is the function which describes the amount of human capital allocated to the production of new human capital over the life cycle, $K(t)$. Clearly the cost of training will be the product $RK(t)$. As such, it is of fundamental importance to establish the empirical validity of the shape of this function as predicted from the model. If the general shape of $K(t)$ is found to be insupportable empirically, the entire model is questionable, since all functions were derived from $K(t)$.

The most complete work to date on empirically estimating on-the-job-training (*OJT*) costs is by Mincer (1962). Using a version of equation (1) which assumed the rates of return were equal across investments, Mincer calculated the additional training costs, both formal and *OJT*, associated with (a), the marginal decision to complete high school if elementary school has been completed, and (b) the marginal decision to complete college if high school has been completed. The procedure used to obtain estimates of these costs year-by-year was first to calculate the internal rate of return to each marginal decision and assume this rate of return represented the rate of return to both schooling and *OJT*.[17]

Table 1 lists all the schooling and *OJT* costs of high school and college graduates over the life cycle for the years 1939, 1949, and 1958. The underline indicates where schooling costs stop and *OJT* costs start, which for high school graduates is age 17, and for college graduates is age 21. The costs listed for high school graduates are the additional training costs incurred by individuals who complete high school over and above those incurred by individuals who have completed elementary school.

[17] Mincer's data were not designed for the sort of test made here, since he was interested in the total marginal cost of the decision rather than year-by-year marginal cost.

TABLE 1—ESTIMATED MARGINAL COSTS OF SCHOOLING AND ON-THE-JOB TRAINING, BY AGE AND LEVEL OF SCHOOLING FOR MALES IN THE UNITED STATES

	High School				College		
Age	1939	1949	1958	Age	1939	1949	1958
14	388	777	1266	14	0	0	0
15	455	939	1538	15	0	0	0
16	545	1121	1917	16	0	0	0
17	643	1309	2338	17	0	0	0
18	254	544	225	18	877	1881	3246
19	200	538	224	19	1096	2268	3776
20	139	441	223	20	1341	2778	4368
21	148	383	222	21	1549	3304	5027
22	158	363	220	22	535	1143	2090
23	170	329	217	23	513	1273	2001
24	169	315	214	24	485	1329	1902
25	168	307	211	25	454	1335	1891
26	167	293	208	26	420	1311	1880
27	166	268	204	27	382	1294	1660
28	159	264	200	28	350	1267	1528
29	151	255	195	29	314	1260	1367
30	142	225	189	30	274	1252	1197
31	131	196	183	31	230	1218	1149
32	118	148	175	32	181	1150	1096
33	105	161	165	33	163	1075	1037
34	90	154	154	34	143	1008	971
35	73	167	141	35	121	884	898
36	54	151	126	36	96	763	815
37	37	143	109	37	68	599	719
38	14	149	89	38	37	432	616
39	..	156	67	39	3	228	501
40		129	43	40		47	423
41		89	16	41		17	339
42		65	..	42		..	245
43		17	..	43		..	144
44				44			27

Source: Mincer (1962).

The costs listed under college graduates are marginal to the decision to go to college given high school has been completed.

The counterpart to Mincer's estimates of marginal training costs would be the difference $R_1K_1(t)-R_2K_2(t)$ in the model presented here. The terms R_1 and R_2 are the rental rates of a college graduate and a high school graduate, respectively, and $K_1(t)$ and $K_2(t)$ are the respective amounts of human capital used in production over the life cycle. Notice that the above difference measures the additional cost of investment associated with the marginal decision to complete college given high school has been completed. The marginal training costs will be proportional to

$$H(t) = K_1(t) - K_2(t) \tag{22}$$

If $H(t)$ is to predict training costs (up to a factor of proportionality), it should, in general, display properties similar to those displayed by Mincer's data. The shape of $H(t)$ can be mathematically determined on the intervals (t_2^*, t_1^*) and (t_1^*, N) where it is assumed that both groups stop earning at age N. On (t_2^*, t_1^*) the individuals represented by $K_2(t)$ will not be specializing in the production of human capital, whereas the individuals

TABLE 2—SUMMARY OF SIGNS ON FIRST AND SECOND DIFFERENCES BY YEAR AND BY SCHOOLING

	High School				College			
	School		*OJT*		School		*OJT*	
	Number positive	Number negative	Number positive	Number negative	Number positive	Number negative	Number positive	Number negative
First Differences								
1939	3	0	3	18	3	0	0	18
1949	3	0	3	22	3	0	3	17
1958	3	0	0	24	3	0	0	22
Total	9	0	6	64	9	0	3	57
Second Differences								
1939	2	0	3	13	1	1	2	15
1949	2	0	9	15	2	0	7	12
1958	2	0	0	16	2	0	4	17
Total	6	0	12	44	5	1	13	44

represented by $K_1(t)$ are still specializing. On this interval, $H(t)$ is increasing at an increasing rate.

On the interval (t_1^*, N) both groups of individuals have stopped specializing and $H(t)$ declines at an increasing rate until just before year N, $H(t)$ starts declining at a decreasing rate. Since the observed differences would become extremely small toward the end of the life cycle, we would not expect to see $H(t)$ decline at a decreasing rate. As can be seen from Table 1, Mincer did not observe differences past age 44.[18]

Thus, this model predicts the signs on both first and second differences of Mincer's data. In particular it predicts that over the interval (t_2^*, t_1^*), both first and second differences should be positive and over the interval (t_1^*, N), both first and second differences should be negative. In looking at the marginal decision to complete high school, t_2^*, is 14 years of age and t_1^* is 17. When the marginal decision to complete college is under consideration, t_2^* is 17 and t_1^* is 21.

[18] It is this fact that leads Ben-Porath (1970) to think individuals may well produce human capital in a nonneutral fashion. He also used Mincer's data.

All the signs of first and second differences are summarized in Table 2. Over the interval (t_2^*, t_1^*), we have a total of 18 first differences taking both high school and college together. All were correctly predicted. On the interval (t_1^*, N) or what is entitled *OJT*, we have a total of 130 first differences of which 121 were correctly predicted.

Again taking both high school and college together, we have a total of 12 second differences on (t_2^*, t_1^*) or what is labeled school. All but one of these were correctly predicted. On (t_1^*, N) there are 113 second differences of which 88 were correctly predicted. Taken individually, in 1949 the null hypothesis that the signs of second differences on (t_1^*, N) were random could not be rejected by a standard chi-square test. In all other cases this null hypothesis was rejected.

IV. Variations in the Period of Specialization

One of the important numbers produced by this model from a given combination of parameter values is the time, t^*, when an individual stops specializing in the production of human capital and begins renting

part of his stock in the labor market. Traditionally, empirical studies have taken this number to be the time when an individual completes formal schooling. In this model, however, t* is interpreted as being the point in time in the investor's life cycle when he ceases to devote 100 percent of his time to the production of human capital. One observation that can be made from the model is that when specialization stops, training does not stop but continues at a decreasing rate. An individual who does not find it optimal to specialize, even if he is in school, will search for work which can accommodate his investment activities and yield income. The question then becomes one of who will shorten the period of specialization by beginning to participate in the labor force.

To get at this question in a way which will test the model we must make some assumptions. The groups used to make the test are white and nonwhite teenagers between 14 and 19 who are attending high school.[19] The assumptions made are that the rate at which human capital depreciates, the rate at which future income is discounted, and the ability to produce is the same on average across any race, sex, or age stratification of teenagers attending school. In other words, in our model, r, δ, and β are assumed to be about equal across any race, sex, or age stratification on average.[20]

With this assumption, the distinctive feature between or within stratifications will be differences in the initial endowment of human capital. One's initial endowment is a single number which summarizes a number of principle components. It is argued here that two such principle components which are very important are family income and family size. Surely there are other environmental effects which are important, but our attention will be directed to these two for the test which follows.

The individual's initial stock, in a sense, determines his relative protagonism toward his stock of human capital. The more conducive is his environment to producing human capital, the larger will be E_0. Family income influences several variables which might create an environment relatively helpful in the production of human capital. Such things as space, privacy, cultural activities, and relatively well-educated parents would be more likely to be present in relatively high income families, and would be associated with a more conducive environment to produce human capital.

A relatively large family, on the other hand, would tend to compete with many of these things as well as with the time parents could spend giving an individual personal attention. Family size would, therefore, tend to have a negative influence on E_0. It seems, however, that since family income can substitute for many of the negative effects of family size, income would be the more reliable indicator of the size of E_0. Certainly it would be expected that the higher the family income, the less important the family size.

The model proposed here predicts that other things constant, the larger one's initial endowment of human capital, the sooner will he stop specializing in the production of human capital. Translated into the variables described above, the larger one's family income is or the smaller one's family size is, the sooner he will begin participating in the labor force.

Since black families tend to be larger than white families and have smaller incomes, we would expect to see a smaller percentage of black teenagers who are attending school participating in the labor

[19] Notice that dropouts are not included.

[20] The existence of market imperfections in the investment market may create problems with this assumption if they are directed at a particular subset of the population. The empirical evidence is not clear cut on this point.

TABLE 3—THE AVERAGE LABOR FORCE PARTICIPATION RATE OF TEENAGERS, ENROLLED IN SCHOOL, BY AGE, COLOR, AND SEX

	Whites		Nonwhites	
Age	Male	Female	Male	Female
14	13.425	5.234	7.576	3.112
15	17.314	7.398	10.550	4.079
16	27.016	14.794	16.087	6.887
17	37.840	24.747	21.663	12.325
18	43.797	33.476	33.870	23.324
19	48.408	41.545	40.547	31.713

Source: Fearn.

force at any given age.[21] Table 3 shows data taken from Robert Fearn which does tend to confirm our hypothesis.

However, this test is not a particularly strong one because the rate of unemployment among black teenagers is relatively high. It may be that the high rate of unemployment forces black teenage students to continue specialization involuntarily. To get around the problem we can look at Fearn's regressions which adjust for unemployment and wage rates as well as family size and income. We can no longer make across strata comparisons, but it still should be true that within a particular age, sex, race category, income should be positively related and family size negatively related to labor force participation.

Table 4 summarizes these regressions and we find that our hypothesis is still true among nonwhites, but not among whites. As pointed out above, it is suspected that family income will become more important than family size as income increases. Further, it is also suspected that at constant family income, as family size increases its effect will become more important. Given that nonwhites have larger

TABLE 4—THE SIGNS OF REGRESSION COEFFICIENTS ON FAMILY INCOME (*FI*) AND FAMILY SIZE (*FS*) BY COLOR, SEX, AND AGE

	White				Nonwhite			
	Male		Female		Male		Female	
Age	*FI*	*FS*	*FI*	*FS*	*FI*	*FS*	*FI*	*FS*
14[a]	+	+	+	+	+	−	+	−
15	+	+	+	+	+	−	+	−
16	+	+	+	+	−	+	+	−
17	−	+	+	+	+	−	+	−
18	+	+	+	+	+	−	−	−
19	+	+	+	+	+	+	+	−

Source: Fearn.
[a] Labor force participation rate is the dependent variable.

families and smaller incomes than whites, it is to be expected that family size would be more important in determining the length of the period of specialization of nonwhites. For this reason, it would seem that the negative effect of family size on labor force participation rates among nonwhites is not spurious.

V. Conclusions

At the end of Section I in this paper, a number of questions were asked about the shape of $k(t)$ over the life cycle if the individual chose $k(t)$ within some choice theoretic framework as it would appear he must do. From the model presented in Section II, a particular $k(t)$ is implied which does not demonstrate all the characteristics shown in Figure 2. Although $k(t)$ is equal to one in the beginning of the life cycle and declines after specialization stops, the resemblance ends there. The function describing the proportion of human capital allocated to production in the model presented here neither has a notch at the time specialization stops nor does it decline at a constant rate. Rather, there is a fairly smooth transition from specialization to nonspecialization and during the period of nonspecialization, $k(t)$ declines first at an increasing rate, then at a decreasing rate.

[21] The average white family had 1.27 children under age 18 whereas the average Negro family had 1.80, in 1969. Only 8.9 percent of white families had a female head whereas 28.6 percent of Negro families were headed by a female. See U.S. Bureau of Census (Table 45, p. 38).

The shape of $k(t)$ is determined by the way which the individual allocates his stock of human capital between production and earning. This process is described by the function $K(t)$. Empirical evidence indicates that the shape of training costs marginal to the next best alternative are predicted reasonably well. Since these marginal costs depend on the shape of $K(t)$ quite directly, possibly the model describes the process at least in gross sense.

A second test was carried out on the important question of how investors react to a change in the initial endowment of human capital as measured by the proxy variables of family income and family size. The results indicate that investors do tend to spend less time specializing in production as the initial stock gets larger, other things the same. This result indicates that taking the time when formal schooling ends as the time when specialization stops may not be entirely accurate in the case of high school and college graduates.[22]

REFERENCES

G. A. Akerlof, "Structural Unemployment in a Neoclassical Framework," *J. Polit. Econ.*, May/June 1969, *77*, 399–407.

K. J. Arrow, "Applications of Control Theory to Economic Growth," in G. Danzig, ed., *Mathematics for the Decision Sciences, Part 2*, Providence 1968, 85–119.

G. Becker, *Human Capital*, New York 1964.

——— and B. Chiswick, "Education and the Distribution of Earnings," *Amer. Econ. Rev. Proc.*, May 1966, *56*, 350–69.

Y. Ben-Porath, "The Production of Human Capital and the Life Cycle of Earnings," *J. Polit. Econ.* July/Aug. 1967, *75*, 352–65.

———, "The Production of Human Capital Over Time," in W. L. Hansen, ed., *Education, Income, and Human Capital*, New York 1970, 129–46.

B. Chiswick, "Human Capital and the Distribution of Personal Income," doctoral dissertation, Columbia Univ. 1967.

R. Fearn, "Labor Force and School Participation of Teenagers," doctoral dissertation, Univ Chicago, 1968.

W. Haley, "Human Capital Accumulation Over the Life Cycle," doctoral dissertation, N. Carolina State Univ. 1971.

T. Johnson, "Returns from Investment in Human Capital," *Amer. Econ. Rev.*, Sept. 1970, *60*, 546–60.

J. R. Millar, "A Theory of On-the-Job Training," papers presented at winter meetings of Econometric Society 1971.

J. Mincer, "On-the-Job Training Costs, Returns, and Some Implications," *J. Polit. Econ.* Oct. 1962, *70*, 50–73.

———, "The Distribution of Labor Incomes: A Survey with Special References to the Human Capital Approach," *J. Econ. Lit.*, Mar. 1970, *8*, 1–26.

H. Oniki, "On Deriving the Individual's Demand Function for Educational Investment," Harvard Institute Econ. Res., disc. paper no. 154, Cambridge, Mass. 1970.

L. S. Pontryagin et al., *The Mathematical Theory of Optimal Processes*, New York 1964.

S. Rosen, "Knowledge, Obsolescence, and Income," Univ. Rochester and Nat. Bur. Econ. Res., unpublished Nov. 1970.

T. W. Schultz, *Investment in Human Capital*, New York 1971.

E. Sheshinski, "On the Individual's Lifetime Allocation Between Education and Work," *Metroeconomica*, Jan. 1968, *20*, 42–49.

R. M. Solow, "Investment and Technical Progress," in K. J. Arrow et al., eds., *Mathematical Methods in the Social Sciences*, Stanford 1959.

Y. Weiss, (1971a) "Ability and Investment in Schooling: A Theoretical Note on J. Mincer's 'Distribution of Labor Income'," *J. Econ. Lit.*, June 1971, *9*, 459–61.

———, (1971b) "Invesment in Graduate Education," *Amer. Econ. Rev.*, Dec. 1971, *61*, 833–52.

U.S. Bureau of Census, *Statistical Abstract of the United States*, Washington 1970.

[22] See T. Schultz, ch. 7, for more evidence on this question.

Inherent Advantage, Costs, and Resource Allocation in the Transportation Industry

By ROSALIND S. SENECA*

The interpretation of the term "inherent advantage" when applied to competing modes of transportation has been the subject of some controversy. It first appears in the Declaration of National Transportation Policy (*NTP*) in the Transportation Act of 1940 which states that the regulatory authorities (Congress, the Commissions, and the Courts) should ensure that traffic is allocated between modes in such a way as to "recognize and preserve the inherent advantages of each," and should also prevent "destructive competition" between modes. The terms inherent advantage and destructive competition are not defined in the Declaration. However, in an amendment to the rules of rate making which appears as Section 15 a(3), the Interstate Commerce Act states that "rates of a carrier shall not be held up to a particular level to protect the traffic of any other mode of transportation, giving due consideration to the objectives of the National Transportation Policy."

A study of the interpretation of the law by the Interstate Commerce Commission (*ICC*) reveals that the *ICC* has tended to interpret inherent advantage to mean lowest average cost, and has sometimes prevented modes with low marginal costs from lowering prices to the marginal cost floor when this would divert traffic from the mode with the lowest average cost (see Robert Harbeson). The *ICC* has also accepted the argument that in determining the price to be charged by any one mode, the costs of that mode alone should be referred to, the costs of the competing modes being irrelevant.

Economists define destructive competition as occurring only when one mode attempts to capture the traffic of a competitor by pricing below marginal cost, and some maintain that, being an irrational market strategy, destructive competition is unlikely to occur (see for example John McGee).[1] However, the *ICC* has evidently been concerned with preventing the demise of any existing mode, and the tendency has been to interpret as destructive competition any price policy of a competitor which diverts traffic away from the mode with the lowest average cost.[2] Indeed, the equity issue of the survival of competing modes has played an important part in the decisions of the regulatory authorities (see Ronald Coase).

In the theoretical literature the question of the determination of the optimal allocation of traffic between transportation modes has been dominated by the marginal cost pricing controversy. A policy of setting rates equal to marginal costs has been

* Assistant professor of economics, Hunter College of the City University of New York. This paper is taken from a dissertation completed at the University of Pennsylvania. I should like to thank O. E. Williamson, C. Caton, J. J. Seneca, G. D. N. Worswick, P. Davidson, D. F. Bradford, and the referee for helpful comments.

[1] Sydney Weintraub, p. 193, presents contrary argument.

[2] An important recent example is the Ingot Molds case in which the Supreme Court decided *not* to allow the railroads to reduce their rates for carrying Ingot Molds to equal that of the barges, a strategy which would enable the railroads to compete for the traffic. The barge average cost was below that of the railroads, but the railroads had the lower marginal cost.

supported on the grounds that it possesses attractive efficiency properties (restated in a recent article by William Baumol and David Bradford). The main objections to such a policy have been 1) that second best considerations destroy its theoretical justification; 2) that where decreasing costs prevail and prices are set equal to marginal costs total costs will not be covered; and 3) that the difficulty of estimating marginal cost makes the practical operation of a marginal cost pricing policy infeasible.

An examination of the testimony presented by several leading economists in the current *ICC* investigation into "Cost Standards in Intermodal Rate Proceedings" reveals continuing differences on this subject. While there is agreement that the function of regulation under the concept of inherent advantage is to permit an allocation of traffic between modes which minimizes the total cost of providing transportation service (indeed "safe, adequate, *economical* and *efficient* service" was a stated goal in the *NTP*) separate arguments are made in favor of allocating traffic 1) to the mode with the lowest "long-run variable cost" (Baumol, 1970); 2) to the mode with the lowest "long-run incremental cost" (Leon Moses); 3) to the mode with the lowest "out-of-pocket cost" (Merton Peck). Peck defines out-of-pocket cost as long-run variable cost, but the term has also been defined elsewhere as meaning "short-run marginal cost" (George Borts).

However Borts alone makes explicit the following important point: namely that there is a serious deficiency in the use of marginal costs to allocate traffic among competing modes, a deficiency which arises because of the failure to take account of differences between the total costs of competing modes. Allocating traffic to the mode with the lowest long-run marginal cost will not *necessarily* minimize the total cost of providing the service.

The purpose of this paper is to provide a further clarification of these issues. In Section I a model is presented for determining the optimal allocation of traffic between two competing modes, where the objective is to minimize the total costs of transporting a fixed amount of a single commodity between two points. The role and importance of total and marginal costs in attaining a solution to the cost-minimization problem is identified. An efficient allocation of resources can only be found by considering the transportation industry as a whole, the cost curves for all modes and the total demand for transportation, and *not* by looking at cost and demand relationships for each mode separately. In Section II an extension of the model is indicated for consideration of the question of optimal price policy when demand for transportation is not perfectly inelastic but represented as a downward-sloping schedule, and also when a profit constraint is imposed for the industry. Section III contains the conclusions which can be drawn from the models.

I

Model 1

Assume that the transportation industry consists of two modes of transportation both operating between the same two geographic points. Assume further that the cost characteristics of a transportation mode can be put in one of two categories:[3]

Category I: The mode operates with short-lived facilities and highly flexible organization. This means all short-run fixed costs are of a transitory nature and can be transformed quickly into incremental costs. Long-run marginal costs include incremental capacity costs and tend to be increasing over the relevant range. (The

[3] See for example Baumol et al. for a fuller analysis of transportation costs.

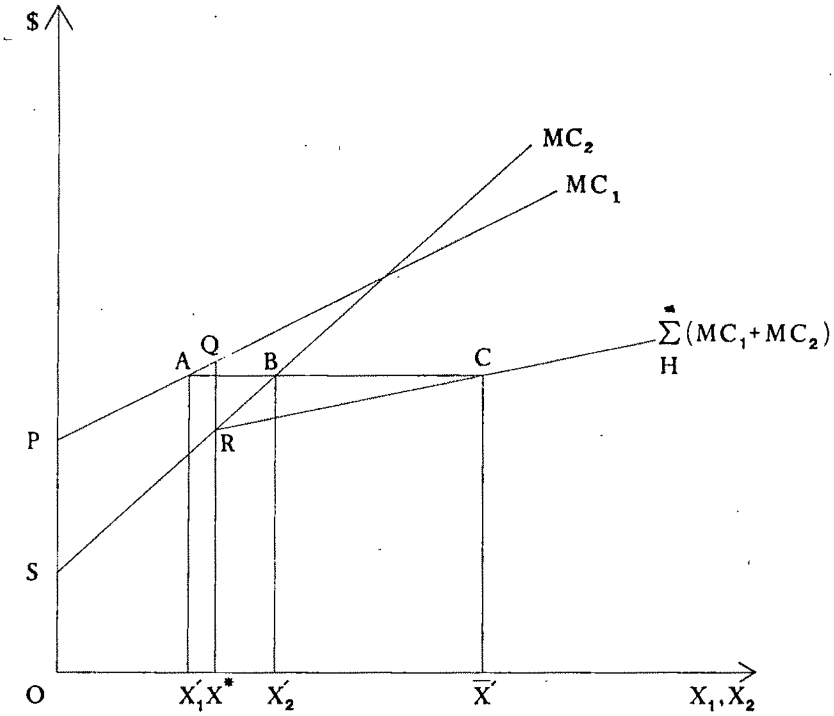


FIGURE 1

trucking industry might fall into this category.)

Category II: The technology of the mode is characterized by large amounts of sunk physical investments in long-lived facilities which are not generally transferable to other uses without great loss. This is typical of the railroads. The result is that, while some capacity costs (e.g. depreciation, cost of capital) should be included in long-run marginal costs, an indefinitely long-term view of incremental costs is not appropriate, and a large proportion of fixed costs may be expected to remain fixed over any contemplated planning horizon. Long-run marginal costs therefore tend to be decreasing over the relevant output range.[4]

[4] There is disagreement over the question of whether or not the railroads in fact operate on the declining portion of their long-run marginal cost curve. Moses and J. R. Meyer et al. suggest that the railroads do operate under increasing returns. Borts disputes this.

Suppose now that there is one commodity, X, to be carried and the demand for X is unresponsive to price changes and fixed at a given level $\bar{X}$. The objective is to minimize the total cost of transporting $\bar{X}$ between two points and to determine how much of $\bar{X}$ should optimally be carried by mode 1 and how much by mode 2.[5]

Case 1: Consider first the case in which both modes fall into Category I. Then the long-run marginal cost curves of mode 1 and mode 2 can be represented by the increasing curves MC_1 and MC_2 in Figure 1. Suppose $\bar{X} < X^*$ then the *total* cost of loading $\bar{X}$ onto mode 2, shown by area $OSRX^*$

[5] A formal presentation of the following models is made in the Appendix to this paper in which necessary and sufficient conditions for cost minimization are given for those cases presented in the text.

is less than the total cost of loading $\overline{X}$ onto mode 1, shown by area $OPQX^*$.[6] All of $\overline{X}$ should be carried by mode 2, mode 1 is not activated. Note also that for $\overline{X}<X^*$, $MC_2<MC_1$. If $\overline{X}$ exceeds X^*, the total transportation cost can be minimized by activating mode 1 and sharing the commodity between the modes. Now the share of $\overline{X}$ to be carried by each mode is determined such that MC_1 and MC_2 are equal, and is found in the following way. Starting from point R the marginal cost curves MC_1 and MC_2 are added horizontally to derive the curve labelled $\sum_H (MC_1+MC_2)$ in Figure 1, where $\sum_H$ denotes "horizontal sum." This summed MC curve measures the opportunity cost of carrying each level of $\overline{X}$. For example, where $\overline{X}$ equals $\overline{X}'$ the opportunity cost of carrying $\overline{X}'$ is given by C on the summed curve. At the level C the marginal costs for each mode are equal, at A for mode 1 and at B for mode 2. The share of $\overline{X}'$ carried by mode 1 is X_1' and the share carried by mode 2 is X_2'. Note that for different levels of demand the optimal division of $\overline{X}$ between the two modes changes. Whether one or both modes should be used does not therefore depend wholly on their cost curves but also on the scale of demand.

Case 2: Consider now a second case in which both modes fall into Category II. Both modes have decreasing long-run marginal cost curves, MC_1 and MC_2 in Figure 2. Suppose $\overline{X}$ is to be carried. At $\overline{X}$, MC_2 lies below MC_1. However this does not necessarily mean that $\overline{X}$ should all be carried on mode 2, since, for lower values of X, a portion of the MC_2 curve lies above the MC_1 curve. Only a comparison of the areas under the curves, i.e., of total costs, can determine whether $\overline{X}$ should be carried on

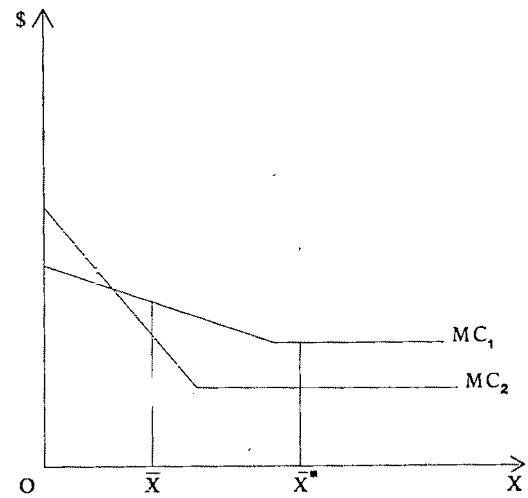


FIGURE 2

mode 1 or mode 2. Also in this case the commodity will never be divided between the modes, regardless of the value of $\overline{X}$, but will always be carried by one mode alone. For example at $\overline{X}$ in Figure 2, TC_1 is less than TC_2 (where TC=total cost), and so all of $\overline{X}$ will be carried by mode 1. When $\overline{X}^*$ is to be carried, TC_2 is just less than TC_1 and so all the commodity will be carried by mode 2. Because both cost curves are decreasing, it is never efficient to shift small amounts of the commodity between the modes since the lower costs can only be taken advantage of if large amounts are carried. The opposite of this is true in Case 1 where small amounts can be carried at lower cost than large amounts.

Case 3: Assume now that mode 1 has an increasing long-run marginal cost curve and mode 2 a long-run marginal cost curve which decreases and asymptotically approaches a constant level above zero. Assume further that MC_2 decreases at a slower rate than MC_1 increases (see Figure 3). Suppose the fixed demand is relatively small. If $\overline{X}\leq X_2$ then $TC_1<TC_2$, since $A_2<A_1$. When mode 1 carries X_2 its marginal cost is just equal by construction to

[6] Since all relevant costs are considered variable in the long run, the area under the marginal cost curve at any output level is a correct measure of the *total* cost of transporting that level of output.

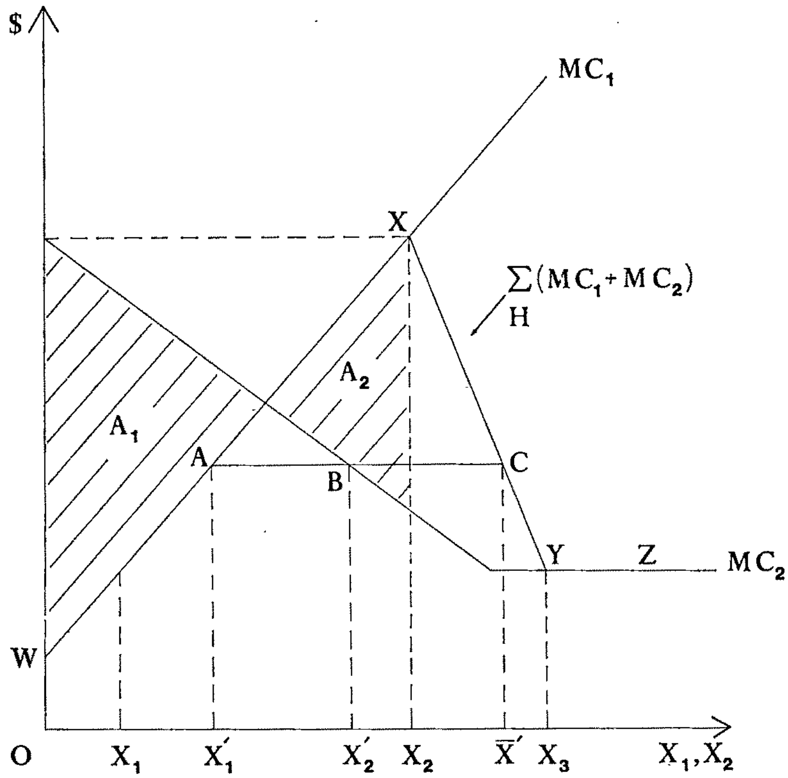


FIGURE 3

the marginal cost on mode 2 at the vertical axis. Therefore, for any $\overline{X}>X_2$, the total transportation cost can be minimized by activating the second mode and taking advantage of the fact that its marginal cost decreases by allocating to it an increasing share of $\overline{X}$. Note that in this case it is cost minimizing to start sharing the commodity between the modes at a *lower* value of $\overline{X}$ than that at which $TC_1=TC_2$, i.e., *before* $\overline{X}$ reaches that level at which $A_1=A_2$.

Now the share of $\overline{X}$ to be carried by each mode is determined such that MC_1 and MC_2 are equal and is found by adding the curves MC_1 and MC_2 horizontally, starting from point X, to derive the forward sloping curve labelled $\sum_H (MC_1+MC_2)$ in Figure 3. As before, this summed MC curve measures the opportunity cost of carrying each level of $\overline{X}$. If $\overline{X}=\overline{X}'$ in Figure 3 the opportunity cost of carrying $\overline{X}'$ is given by C on the summed curve, and the share of $\overline{X}'$ to be carried by modes 1 and 2 equals X_1' and X_2', respectively, since these amounts can be carried at equal marginal cost. Further inspection of Figure 3 shows that the nearer the value $\overline{X}$ is to X_3 the larger the share of commodity loaded onto mode 2, the smaller the share loaded onto mode 1 and the lower the marginal cost of transporting the commodity. If $\overline{X}$ exceeds X_3 then only X_1 is loaded onto mode 1 and all the rest onto mode 2, since additional quantities of commodity can be loaded onto mode 2 at constant marginal cost. The composite curve $WXYZ$ therefore measures the marginal cost of carrying the commodity at every level of $\overline{X}$.

For the case in which the marginal cost

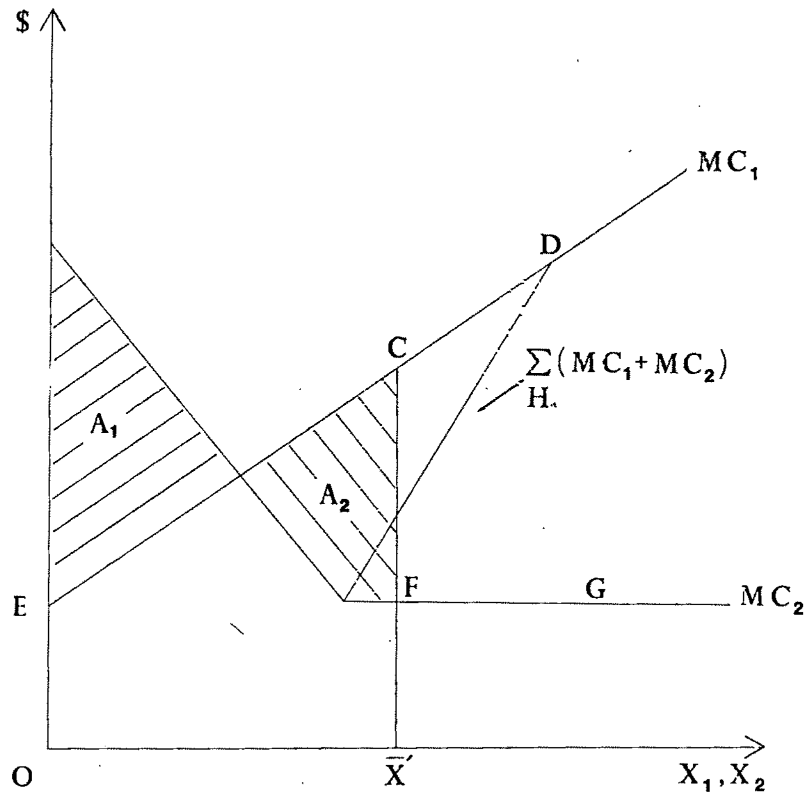


FIGURE 4

curve for mode 1 increases at a slower rate than the marginal cost curve for mode 2 decreases the summed MC curve is backward sloping (see Figure 4). For low values of $\overline{X}$ the total transportation cost is minimized by using mode 1 alone. However there is a value for $\overline{X}$, say $\overline{X}'$, at which the area under the MC_1 curve is precisely equal to the area under the MC_2 curve, i.e., $A_1 = A_2$. At higher output levels $A_2 > A_1$ and hence at $\overline{X}'$ all of the commodity should be shifted onto mode 2. This point occurs *before* the marginal cost of carrying one extra unit of commodity by mode 1 equals the marginal cost of carrying the first unit of commodity by mode 2, point D in Figure 4. Further we can say that at no point is the backward-bending part of the summed MC curve used to determine the amounts of commodity carried by each mode. The marginal cost of transporting each level of X is shown by the discontinuous curve $ECFG$, the discontinuity occurring as mode 2 is brought into operation at $\overline{X}'$.

II

Model 2

In this section the model is extended: the assumption that the demand for transportation is fixed and unresponsive to price changes is dropped and the model redefined using a downward sloping demand schedule for transportation.

Assume that the transportation industry consists of two modes with the same cost characteristics as described in Case 3 above. A single firm's demand curve for transportation is derived from the demand for the commodity produced by the firm.

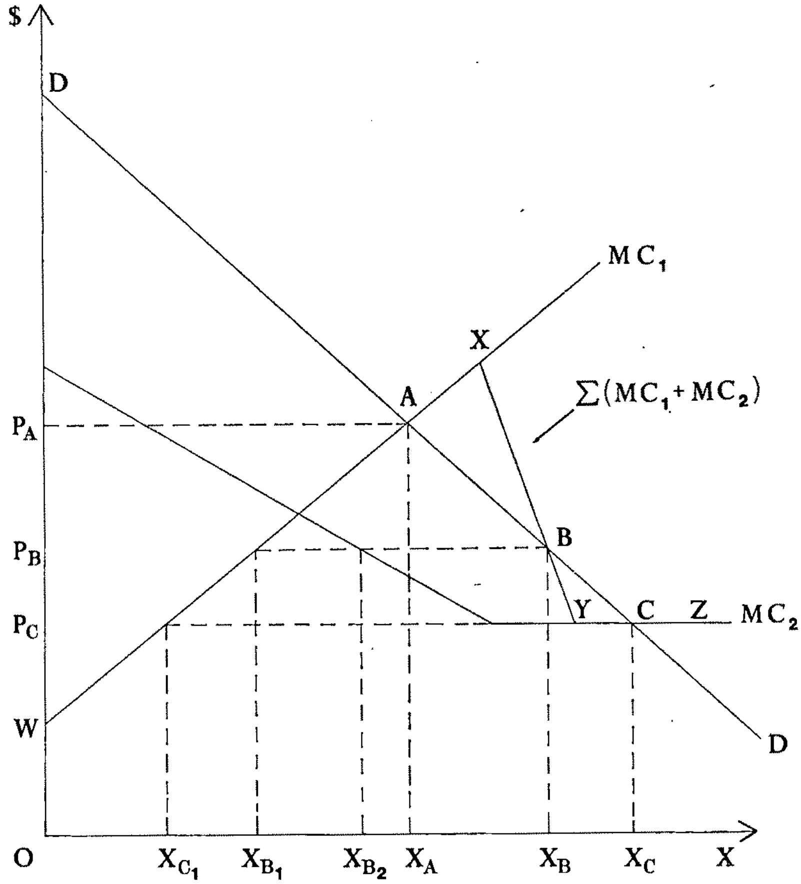


FIGURE 5

A firm requires transportation for the commodity between two points and this transportation requirement therefore appears as an input in the firm's production function. From the profit-maximizing calculation of the firm we can derive its demand curve for transportation, and, assuming perfect competition, we can express the total demand for transportation by all firms in an industry as $P=G(X)$, where $P=$price of transportation and $X=$the number of units of commodity to be transported.[7] The demand curve is shown as *DD* in Figure 5.

[7] Note that although there are two modes the demand for transportation depends on a *single* transportation price. The underlying assumption is that industry considers the two modes to be perfect substitutes and so will always choose the lowest priced mode.

Now the objective is to maximize the social welfare derived from the productive activity of the transportation industry. We shall use the conventional partial equilibrium social welfare function whereby if it is assumed that all of the optimum conditions of production and exchange are satisfied elsewhere in the economy, social benefit (*SB*) is given by total revenue (*TR*) plus consumers' surplus (*CS*), and social cost (*SC*) is separable into total pecuniary cost less intramarginal rent. Assuming that all factors are available to each mode in completely elastic supply, intramarginal rents will be zero and the net welfare gain is:

$$W = SB - SC = TR + S - TC$$

The necessary conditions for maximiz-

ing W are found to be $P=MC_1=MC_2$, i.e., the total amount of commodity carried, and the division of the commodity between each mode are determined such that price equals the marginal cost of carrying the commodity on each mode.[8]

Using the cost curves MC_1 and MC_2 described in Figure 3 we know that the composite curve $WXYZ$ measures the marginal opportunity cost of carrying each different amount of commodity. We now introduce the downward-sloping demand schedule DD and observe that it intersects the composite cost curve at three points—A, B, and C in Figure 5. Thus, one of three prices could be charged. If P_A were charged, X_A would be demanded and only mode 1 would be activated. If P_B were charged, X_B would be demanded; mode 1 would carry X_{B1} and mode 2 would carry X_{B2}. If P_C were charged X_C would be demanded; mode 1 would carry X_{C1} and the rest would be carried by mode 2.

Of all three points A, B, and C, that price is chosen which yields the largest value of W. This can *only* be found by comparing the appropriate areas under the demand and cost curves. Once again reference to marginal costs alone is insufficient for determining the optimal price and allocation of commodity between modes, total costs must be referred to in the solution.

Model 3

Suppose now that the condition is made that the total revenues received by the transportation industry as a whole must exceed the total costs of the industry by some fixed amount M, i.e., $TR-TC=M$. The objective is now to maximize social welfare (W) subject to the constraint $TR-TC-M=0$. The necessary conditions for a maximum are found to be:

$$P-\frac{MC_1}{P}=\frac{\lambda}{(1-\lambda)}\,\frac{1}{E}$$

$$P-\frac{MC_2}{P}=\frac{\lambda}{(1-\lambda)}\,\frac{1}{E}$$

where $E=(P/X)(dX/dP)$ is the elasticity of demand for transportation, and λ is a multiplier reflecting the cost or reduction in welfare resulting from the presence of the break-even constraint.[9]

These conditions are identical to those derived by Baumol and Bradford and simply show that in this single commodity model when a profit constraint is present the optimal price should exceed marginal cost by an amount proportional to the inverse of the elasticity of demand. Thus, when the elasticity of demand is high price will exceed marginal cost by less than when the elasticity of demand is low. The conditions also imply that the quantities of commodity carried by each mode must be

[8] Let $C_1(X_1)$ = the long-run total cost function of mode 1
$C_2(X_2)$ = the long-run total cost function of mode 2
where X_1 = amount of commodity carried by mode 1
and X_2 = amount of commodity carried by mode 2

The objective is to maximize

$$W=\int_0^X G(X)dX-C_1(X_1)-C_2(X_2);$$
$$X=X_1+X_2$$

The necessary conditions for a maximum are

$$\frac{\partial W}{\partial X_1}=P-\frac{\partial C_1}{\partial X_1}=0$$
$$\frac{\partial W}{\partial X_2}=P-\frac{\partial C_2}{\partial X_2}=0$$

Hence $P=MC_1=MC_2$

[9] The objective is to maximize $W=\int_0^X G(X)dX-C_1(X_1)-C_2(X_2)$ subject to: $TR-TC-M=PX-C_1(X_1)-C_2(X_2)-M=0$

The maximization calculation yields

$$P-MC_1=\lambda[MR-MC_1]$$
$$P-MC_2=\lambda[MR-MC_2]$$

where MR (marginal revenue) $=P+X(dP/dX)$ substitution of $E=(P/X)(dX/dP)$ into these expressions yields the necessary conditions stated in the text.

such that the marginal costs of each mode are equal.

III. Conclusions

Consider first the problem of cost. From the models presented above it is clear that where a fixed amount of transportation service is to be allocated between two modes, it is total cost and not marginal cost which must be the basis of rational allocation. Moreover any change in total demand means that the optimal share of commodity to be carried by each mode will change, since a change in demand involves a whole new solution to the cost-minimization problem. It is possible that small changes in demand could be dealt with by comparing marginal costs and allocating the extra commodity to the mode with the lowest marginal cost at the point of the previous solution. Also it would be possible to discover a misallocation of resources if differences were observed in marginal costs. However, if there were substantial changes in demand a simple comparison of marginal costs would not lead to the correct solution.

The question of inherent advantage should be discussed in the light of these conclusions. If, by preserving the inherent advantage of each mode, Congress is in fact attempting to ensure that transportation services are provided at minimum cost, then it is the cost structure of the industry as a whole which must be considered and not the cost curves of each mode in isolation. The question of which mode should carry the traffic cannot be answered in simple terms. It makes no sense to suggest that one mode has an inherent advantage and should therefore carry the traffic unless this is the solution obtained from the cost minimization problem *at the specified level of demand*.

It should be noted here that the cost evidence typically used by the *ICC* in making decisions in contested rate cases consists of a single cost figure for each mode (usually fully distributed costs, sometimes marginal cost) calculated at the current output level.[10] The models make it clear that at least some evidence of how costs change with respect to the quantity of transportation provided would be a step towards better decision making.

In Model 2 a downward-sloping demand schedule for transportation was introduced. The maximization of social welfare results in a single optimal price for transportation to be charged by both modes and equal to the marginal cost of each mode. As before, information about total costs is needed to find the optimal solution, and, as before, any change in demand will change the optimal division of commodities between the modes and the optimal price.

Consider now the second best solution to the welfare-maximization problem in which each mode must at least break even, or make a positive profit. Congress has traditionally been concerned with the need to preserve competitive practices in the transportation industry. Both the 1940 Transportation Act and also Section 15 a(3) stress the goal of efficiency and the need to preserve competition to ensure the achievement of the efficiency objective. However, it is clear from recent *ICC* policy that the interpretation of Congressional intent has resulted in an attempt to protect competitors from the inevitable effects of competition, and the refusal to allow price reductions, other-

[10] For example, in the Ingot Molds case the barges declared their fully distributed cost to be \$5.1959 per gross ton. The railroads declared their out-of-pocket cost to be \$4.689 per gross ton and presented several different calculations of fully distributed cost ranging from \$7.80 per gross ton to \$6.613 per gross ton. It is not my intention to comment here on the usefulness of fully distributed cost as a cost measure, merely to point out that unless long-run average cost is constant for every level of output, a single cost figure measured at the current output level is an insufficient guide to the shape of the total cost function.

wise justifiable on efficiency grounds, which threaten at least the profits and possibly the existence of a competing mode. Since the desire to protect competitors often conflicts with the objective of achieving efficiency through competition, a decision by Congress to preserve the existence of competitors cannot be explained on rational economic grounds alone, and must therefore be accepted as part of some social preference function which is revealed in the actions of Congress and the *ICC*. While the economist may not be expected to explain the social preferences behind Congressional action, he can clarify the cost society incurs by attempting to preserve the existence of all competing modes.

In Model 3, the value of the multiplier λ can be interpreted in the following way. Suppose that the initial situation is an ideal one in which all prices in the economy are set equal to marginal costs. The decreasing cost mode is therefore making a loss but is being subsidized by the government using a system of lump sum taxes and transfers which does not distort marginal relationships. When the profit constraint ($TR-TC=M$) is imposed on the industry the price of transportation rises and the quantity provided decreases. The multiplier λ is the dollar value which buyers of transportation would put on a one dollar reduction in the size of M. It is the price which transportation consumers would pay to be able to return to the initial situation. In this sense λ measures the social welfare loss incurred by insisting that the industry as a whole make a profit. Presumably, a change in the profit level required for either or both modes will affect the price and quantity carried by both modes. In particular, an increase in the profit constraint for the increasing cost mode would increase the quasi-optimal price and decrease the quantity carried by all modes (since the demand curve is downward sloping). Moreover it would also tend to reduce the share of traffic carried by the decreasing cost mode, since, when the quantity of traffic to be carried is large, a greater proportion will be carried by the decreasing cost mode, but when the quantity is small the increasing cost mode will carry most of it.

Note that if the initial situation is one in which the existence of decreasing costs has led to subsidization or monopoly pricing which distorts the allocation of resources away from the Pareto optimal allocation then social welfare losses are already being incurred. In this case the introduction of a break-even constraint on the decreasing cost industry might reduce welfare losses. Certainly the conditions for constrained welfare maximization derived by Baumol and Bradford yield a price which is less distorting than that which would be charged by a profit-maximizing monopolist.

Appendix

Formally, the cost-minimization problem can be expressed as follows:

Let $C_1(X_1)=$ the long-run total cost function of mode 1

Let $C_2(X_2)=$ the long-run total cost function of mode 2

where $X_1=$ amount of commodity carried by mode 1

and $X_2=$ amount of commodity carried by mode 2

Then the objective is to minimize

$$C_1(X_1)+C_2(X_2) \tag{1}$$

subject to: $X_1+X_2=\overline{X}$

Let

$$\begin{aligned} f(X_1,\overline{X}) &= C_1(X_1)+C_2(X_2) \\ &= C_1(X_1)+C_2(\overline{X}-X_1) \end{aligned}$$

Now $f(X_1,\overline{X})$ can be expressed as

$$(2) \quad f(X_1, \overline{X}) = \int_0^{X_1} MC_1(\mu)d\mu + \int_0^{\overline{X}-X_1} MC_2(\mu)d\mu$$

Given $\overline{X}$, the conditions for finding that value of X_1, $(\hat{X}_1)$, which minimizes $f(X_1, \overline{X})$ can be derived as follows (note that these conditions will only indicate a *local* minimum).

There are three possible types of solution:

1. If $\hat{X}_1=0$, then we must have

$$(3) \quad f'(\hat{X}_1, \overline{X}) = MC_1(\hat{X}_1) - MC_2(\overline{X} - \hat{X}_1) \geq 0$$

where

$$f'(\hat{X}_1, \overline{X}) = \frac{df}{dX_1}$$

but if $\hat{X}_1=0$ and $f'(\hat{X}_1)=0$ then we must have

$$(4) \quad f''(\hat{X}_1, \overline{X}) = \frac{dMC_1}{dX_1}(\hat{X}_1) + \frac{dMC_2}{dX_1}(\overline{X} - \hat{X}_1) \geq 0$$

where

$$f''(\hat{X}_1, \overline{X}) = \frac{d^2f}{dX_1^2}$$

2. If $0<\hat{X}_1<\overline{X}$, then we must have *both*

$$f'(\hat{X}_1, \overline{X}) = 0 \quad and \quad f''(\hat{X}_1, \overline{X}) \geq 0$$

3. If $\hat{X}_1=\overline{X}$, then we must have $f'(\hat{X}_1, \overline{X}) \leq 0$. If $\hat{X}_1=\overline{X}$ *and* $f'(\hat{X}_1, \overline{X})=0$, then we must have $f''(\hat{X}_1, \overline{X}) \geq 0$.

To illustrate the meaning of these conditions consider the case shown in Figure 3.

Suppose $\overline{X}=\overline{X}'$ in Figure 3. Then it is true that $0<\hat{X}_1<\overline{X}$; in other words, both X_1 and X_2 are positive and both modes are in operation.

From point 2 above the following conditions must be fulfilled:

$$(5) \quad f'(\hat{X}_1, \overline{X}) = MC_1(\hat{X}_1) - MC_2(\overline{X} - \hat{X}_1) = 0$$

and

$$(6) \quad f''(\hat{X}_1, \overline{X}) = \frac{dMC_1}{dX_1}(\hat{X}_1) + \frac{dMC_2}{dX_1}(\overline{X} - \hat{X}_1) \geq 0$$

Condition (5) is nothing more than the equal marginal cost condition. For condition (6) to hold we must have

$$\frac{dMC_1}{dX_1}(\hat{X}_1) \geq \frac{dMC_2}{dX_1}(\overline{X} - \hat{X}_1)$$

since $dMC_2/dX_1(\overline{X}-\hat{X}_1)$ is negative. In other words, the marginal cost curve of mode 1 must be increasing at a faster rate than the marginal cost curve of mode 2 is decreasing; the summed marginal cost curve is forward sloping as shown in Figure 3.

Now consider the case shown in Figure 4. Suppose $\overline{X}<\overline{X}'$ in Figure 4. Then it is true that $\hat{X}_1=\overline{X}$, therefore we must have $f'(\hat{X}_1, \overline{X}) \leq 0$ (from point 3 above), i.e.,

$$MC_1(\hat{X}_1) \leq MC_2(\overline{X} - \hat{X}_1)$$

As Figure 4 is drawn there is no internal solution for $\hat{X}_1$. Therefore *either* $\hat{X}_1=0$, *or* $\hat{X}_1=\overline{X}$, and it will *not* be the case that $f''(\hat{X}_1, \overline{X}) \geq 0$.

Further examination of Figures 1 and 2 will show that these cases can also be interpreted using the above conditions for a local minimum.

REFERENCES

W. J. Baumol, Verified Statement, Appendix A, Railroads' Submission to the ICC, *Docket No. 34013* (Sub. No. 1), Jan. 30, 1970.

——— and D. F. Bradford, "Optimal Departures from Marginal Cost Pricing," *Amer. Econ. Rev.*, June 1970, *3*, 265–83.

——— et al., "The Role of Cost in the Minimum Pricing of Railroad Services," *J. Bus., Univ. Chicago*, Oct. 1962, *35*, 1–10.

G. H. Borts, Verified Statement, *Docket No.*

34013, (Sub. No. 1) *ICC,* Washington, Mar. 29, 1970.

———, Book review of *The Economics of Competition in the Transportation Industries, J. Polit. Econ.,* Apr. 1960, *68,* 205–06.

R. Coase, "The Theory of Public Utility Pricing and its Application," *Bell J. Econ.,* spring 1970, *1,* 113–28.

R. W. Harbeson, "Recent Trends in the Regulation of Intermodal Rate Competition in Transportation," *Land Econ.,* Aug. 1966, *42,* 315–26.

J. McGee, "Predatory Price Cutting: The Standard Oil (N.J.) Case," *J. Law Econ.*. Oct. 1958, *1,* 137–69.

J. R. Meyer et al., *The Economics of Competition in the Transportation Industry,* Cambridge, Mass. 1959.

L. Moses, Verified Statement, Appendix C, Railroads' Submission to the *ICC, Docket No. 34013* (Sub. No. 1), Jan. 30, 1970.

M. J. Peck, Verified Statement, Appendix B, Railroads' Submission to the *ICC, Docket No. 34013* (Sub. No. 1), Jan. 30, 1970.

S. Weintraub, *Intermediate Price Theory,* Philadelphia 1964.

O. E. Williamson, "Peak-Load Pricing and Optimal Capacity Under Indivisibility Constraints," *Amer. Econ. Rev.,* Sept. 1966, *56,* 810–27.

Interstate Commerce Commission, "Cost Standards in Intermodal Rate Proceedings," *Docket No. 34013* (Sub. No. 1), 1970.

Ingot Molds, Pa. to Steelton, Ky., 326 ICC. 77, 33 (1965).

National Transportation Act, 49 U.S.C. Proceeding 1 (1964) (originally enacted as Act of September 18, 1940, ch. 772. 1, 1–54 Stat. 899).

Optimal Restrictions on Foreign Trade and Investment: Note

By RAVEENDRA N. BATRA*

Optimal policy in the presence of international trade and investment, which has been previously analyzed by Murray Kemp and Ronald Jones (1967) in the context of a two-factor, two-good general equilibrium model, constitutes the subject of analysis in a recent article in this *Review* by Franz Gehrels. Gehrels' framework, which postulates a three-factor, two-good economy, differs from that of Kemp and Jones in format but not in substance. Nevertheless, Gehrels has made an important contribution by providing a more complete discussion of "partial" optimization and by deriving optimal tariffs and taxes in the presence of the interindustry wage differential that characterizes factor markets in the underdeveloped countries.

The objective of this note is to show that Gehrels' optimal policy in the presence of the wage differential is incomplete and that his modification of the optimum tariff formula is not always valid.

The assumptions and notations used in this note are the same as those introduced by Gehrels. The problem is to maximize the social utility function

$$(1) \qquad U = U(C_1, C_2) = U(Y_1 - X_1,\ Y_2 - X_2)$$

subject to the production constraint

$$(2) \qquad \phi(Y_1, Y_2, F) = 0$$

and the balance-of-payments constraint

$$(3) \qquad X_1 + X_2\pi + F\rho = 0$$

where C_1 and C_2 are consumption of good 1 and good 2, Y_1 and Y_2 are their outputs, and X_1 and X_2 are their exports or imports depending on whether X_i is positive or negative; F is the amount of foreign investment which is positive if the country in question is a net creditor and negative if it is a net debtor; ρ is the foreign rate of interest; and π is the international terms of trade with the first commodity serving as numeraire.

In conformity with the reality in several developing countries, Gehrels assumes that the country in question is a net debtor, so that $F<0$, imports X_2, so that $X_2<0$, and that the wage differential is paid by the import-competing good, Y_2. Two effects follow immediately from the presence of the differential. First, the transformation curve shifts toward the origin, and second, the marginal rate of transformation no longer reflects the commodity-price ratio, that is,

$$-\frac{\partial Y_1}{\partial Y_2} = bp$$

where $b>1$, indicating that the production costs in Y_2 exceed its social opportunity cost. Let t stand for the rate of tariff; then $p=\pi(1+t)$. The optimum tariff formula derived by Gehrels here is given by his equation (26) which furnishes

$$t = \frac{p-\pi}{\pi} = \frac{X_2}{\pi}\cdot\frac{\partial\pi}{\partial X_2} + \frac{F}{\pi}\cdot\frac{\partial\rho}{\partial X_2} - (b-1)\frac{\partial Y_1}{\partial X_2}$$

As Gehrels shows, $\partial\pi/\partial X_2$ and $\partial\rho/\partial X_2$ are both negative and with X_2 and F also negative, each of the first two terms in this formula is positive. At the start of his paper, he contends that $\partial Y_1/\partial X_2<0$ and $\partial Y_2/\partial X_2>0$, irrespective of the sign of X_i. With X_2 equaling the import, this must be taken to mean that a rise in the relative price of the second good leads to a decline in Y_1 (so that $dY_1<0$) and a decline in the demand for imports (so that with $X_2<0$, $dX_2>0$), which eventually

* Professor of economics, Southern Methodist University.

implies that $(\partial Y_1/\partial X_2)<0$. With $(b-1)>0$, all this leads us to Gehrels' result that the optimal tariff is not only positive, but the existence of the differential also causes a rise in the import duty, if welfare is to be maximized.

However, in the presence of the wage differential, the sign of $(\partial Y_1/\partial X_2)$ is no longer certain. Working with a two-factor, two-good model, Jagdish Bhagwati and T. N. Srinivasan, Jones (1971), and Batra and Prasanta Pattanaik among others have shown that the output of a commodity may actually decline in response to a rise in its relative price when factor markets are distorted by the existence of a wage differential. Furthermore, this result is independent of the shape of the transformation curve. There is no reason why this result should not continue to hold in the three-factor, two-good model, the one used by Gehrels. Thus $\partial Y_1/\partial X_2$ may be positive or negative. If $\partial Y_1/\partial X_2$ is positive, we reach the surprising conclusion that the optimal tariff in the presence of the distortion will be less than what it would be in its absence.

Of greater interest is the fact that the nature of the price-output response in the presence of the differential makes no difference to the optimal tax (or subsidy) on foreign investment. This becomes clear from an examination of the terms contained in Gehrels' equation (32), where[1]

$$\frac{r-\rho}{\rho}=\frac{X_2}{\rho}\frac{\partial\pi}{\partial F}+\frac{F}{\rho}\frac{\partial\rho}{\partial F}+\frac{p}{\rho}\cdot\frac{(b-1)}{b}\frac{\partial Y_2}{\partial F}$$

The sign of $\partial\rho/\partial F$ depends on conditions in the foreign country and if the factor markets there are perfect, $\partial\rho/\partial F$ is unmodified. The signs of $\partial\pi/\partial F$ and $\partial Y_2/\partial F$ are governed by the relationship between factor endowments and the commodity outputs and this relationship is not affected by the existence of the differential.[2] Thus, none of the terms in the optimal tax formula depends on the price-output response, so that Gehrels' results in this case are unscathed.

In the presence of distorted factor markets, a distinction must be introduced between the second best and the first best optimal policy. The optimal tariffs and taxes derived by Gehrels will only lead to the second best optimum, for the inefficiency caused by the existence of the differential in shrinking the transformation curve toward the origin will still remain. The first best policy, the policy which will truly maximize welfare, consists of the grant of factor tax-cum subsidy which has the effect of eliminating the wage differential to the domestic producers of the two outputs (but not to the suppliers of capital or labor), plus the optimal tariff and tax to capture the advantage of the country's monopoly power in international trade and investment. This policy will have the effect of equating b to unity and thus eliminating the last terms from the optimal tariff and tax formulae. At the same time, the economy's production point will move back to somewhere on the undistorted transformation curve.

REFERENCES

R. N. Batra and P. K. Pattanaik, "Factor Market Imperfections, the Terms of Trade, and Welfare," *Amer. Econ. Rev.*, Dec. 1971, *61*, 946–55.

J. Bhagwati and T. N. Srinivasan, "The Theory of Wage Differentials: Production Response and Factor-Price Equalization," *J. Int. Econ.*, Feb. 1971, *1*, 19–35.

F. Gehrels, "Optimal Restrictions on Foreign Trade and Investment," *Amer. Econ. Rev.*, Mar. 1971, *61*, 147–59.

[1] The reading of Gehrels' article is made difficult by the presence of several minor typographical errors. Thus, in the last paragraph of p. 154,

$$\frac{(p-r)}{\pi}\frac{F}{X_2}$$

should actually be

$$\frac{(\rho-r)}{\pi}\frac{\partial F}{\partial X_2};$$

in equation (27) on p. 156, $(\partial Y_1/\partial Y_2)$ should be replaced by $(\partial Y_1/\partial X_2)$; finally in equation (32), $(r-p)$ should read $(r-\rho)$.

[2] See Jones (1971), pp. 440–43.

R. W. Jones, "International Capital Movements and the Theory of Tariffs and Trade," *Quart. J. Econ.*, Feb. 1967, *81*, 1–38.

———, "Distortions in Factor Markets and the General Equilibrium Model of Production," *J. Polit. Econ.*, June 1971, *79*, 437–59.

M. C. Kemp, "The Gain from International Trade and Investment: A Neo-Heckscher-Ohlin Approach," *Amer. Econ. Rev.*, Sept. 1966, *56*, 788–809.

Residential Choice and Air Pollution: A General Equilibrium Model

By EUGENE P. SESKIN*

Pigou's parable of the smoke belching factory imposing an externality on the neighboring laundry has elicited more controversy than one could have expected from such a simple situation. Ronald Coase claimed that the Pigouvian solution of taxes and subsidies was demonstrably inefficient while William Baumol recently defended Pigou by presenting a situation in which a tax placed upon the factory (without taxation or compensation to the laundry) optimizes resource allocation under pure competition.

What if the externality affects the factory itself? Suppose the smoke reduces the work efficiency of employees and causes ill health. It might be conjectured that nothing new is gained by adding the wrinkle of poisonous air which sickens factory workers, since the factory both generates the smoke and is affected by it. By analogy, a "smoke city" where everyone works in factories generating air pollution (which affects only the city) might be deemed consistent with Pareto optimality, presuming the factories compensate individuals for the air pollution with higher wages.[1] Similarly, recent policy discussions have advocated "exporting" air pollution by importing pollution-producing goods. Little concern has been given to the welfare implications since the "republic of smoke" would be experiencing the profit as well as the pollution from production. There are even proponents of the view that some regions of the United States ought to be kept quite clean, while others are allowed to become highly polluted. Their reasoning is that if workers are mobile and acting in a competitive market, they will demand and receive higher wages for living in the polluted locality and the resulting situation will be Pareto optimal.

This paper will show these arguments to be fallacious. For externalities such as air pollution, it is precisely the assumptions of competition and mobility which destroy Pareto optimality. Restoration of efficiency can occur only through outside intervention designed to alter factor prices.

I. The Model[2]

Suppose there exists an economy comprised of two cities: one producing a good which does not cause air pollution (we shall name it bread and refer to the area as bread city); the other manufacturing a commodity which produces air pollution as a by-product (we shall name it steel and refer to the region as steel city). Further, we assume a population of individuals (normalized so that the total population is equal to 1)[3] allocated between the two cities with θ living and working in steel city and $1-\theta$ inhabiting bread city.[4] The problem is determining whether with free movement between the cities (no transportation or moving costs), the wage rates will adjust so that the total population will allocate itself in a Pareto optimal fashion.[5]

* The Urban Institute. I am deeply indebted to Nancy M. Gordon, Lester B. Lave, Robert E. Lucas, Jr., and Thomas E. Morton for helpful comments and suggestions. I am especially indebted to Herbert A. Simon, who is co-author of the Appendix section.

[1] This assumes competition and mobility among workers.

[2] The basic framework of this model is similar to the transportation parables of Robert Strotz.

[3] Note, in checking the units of the various expressions we will derive, this normalization results in an occasional implicit multiplicative factor of 1 which has the units of population.

[4] An alternative interpretation is to regard the total population as allocating θ percent of their time to living and working in steel city and $1-\theta$ percent of their time to living and working in bread city.

[5] In our model the particular Pareto optimum will correspond to the maximum of a linear social welfare function, see below.

A. *The Utility Function*

Each individual in the population is assumed to have an identical utility function.[6] Since conditions in the two cities differ (air pollution is not present in bread city), we shall represent the utility for a resident of steel city as shown in equation (1):

$$u^s = u(s^s, b^s, a) \tag{1}$$

where s^s is the steel consumption per capita in steel city, b^s is the per capita bread consumption in steel city, and a is the amount of air pollution to which an individual in steel city is subjected.[7] Similarly, utility for each bread city resident is:

$$u^b = u(s^b, b^b, 0) \tag{2}$$

The amount of bread and steel consumed by the individual enters each utility function positively, u_1 and $u_2 > 0$ (where the subscripts refer to the partial derivatives with respect to that argument). The partial derivative with respect to a, u_3, is negative, indicating that air pollution is a "bad."

B. *Production Possibilities*

The only resource available for production of bread and steel is labor. Thus, there is a fixed aggregate quantity, 1 (total population), of this homogeneous productive resource available to the economy. The distribution of population will uniquely determine the production of bread and steel. Both commodities are produced in perfectly competitive industries under constant returns with each individual producing one unit of the relevant commodity, i.e., each bread worker produces one unit of bread.[8] Total bread output is equal to the population of bread city, $1-\theta$, and is distributed such that θb^s represents the population of steel city multiplied by the bread consumption per capita for steel city residents and $(1-\theta)b^b$ is the population of bread city multiplied by the bread consumption per capita for bread city residents. An analogous interpretation applies to steel production. Thus, the production constraints in per capita terms are:[9]

$$\theta b^s + (1-\theta)b^b = 1-\theta \text{ (bread)} \tag{3}$$

$$\theta s^s + (1-\theta)s^b = \theta \text{ (steel)} \tag{4}$$

For simplicity we assume that each steel worker not only produces a unit of steel, but also produces a unit of pollution in the making of steel ($a \equiv \theta$); henceforth θ will replace a as the value of the third argument in the utility function for the people of steel city. Thus, the number of units of pollution is identical to the number of units of steel, which in turn is equal to the population of steel city.

II. Individual Utility Maximization

Let each individual be paid the value of his marginal product, i.e., the steel worker's income is $1 \text{x} p^s$ (one unit multiplied by the price of steel) and the bread worker's income is $1 \text{x} p^b$ (one unit multiplied by the price of bread). The consumer is then faced with maximizing his utility subject to his budget constraint. We formulate the respective Lagrangeans:

$$\begin{aligned} &u(s^s, b^s, \theta) \\ &\quad + K^s(p^s - p^s s^s - p^b b^s) \qquad \text{(steel city)} \end{aligned} \tag{5}$$

$$\begin{aligned} &u(s^b, b^b, 0) \\ &\quad + K^b(p^b - p^s s^b - p^b b^b) \qquad \text{(bread city)} \end{aligned} \tag{6}$$

where K^s and K^b are Lagrange multipliers. We then take the partial derivatives with respect to steel and bread consumption, the variables at the consumer's disposal:[10]

$$u_1^s - K^s p^s = 0 \tag{7}$$

[6] Any two individuals consuming the same bundle of goods (including air pollution) would derive equal levels of satisfaction.

[7] There is no superscript on a since air pollution is not present in bread city. Further, it is assumed that air pollution is uniform throughout steel city with each individual consuming the same amount; it is a public "bad."

[8] One can think of individuals owning shares in the bread and steel factories.

[9] Note, the equalities in equations (3) and (4) imply full utilization of the goods.

[10] We assume the individual in steel city regards θ as a parameter, i.e., the level of pollution is independent of his steel consumption.

$$u_2^s - K^s p^b = 0 \tag{8}$$

$$u_1^b - K^b p^s = 0 \tag{9}$$

$$u_2^b - K^b p^b = 0 \tag{10}$$

Dividing (7) by (8) and (9) by (10) we obtain:[11]

$$\frac{\bar{u}_1^s}{\bar{u}_2^s} = \frac{\bar{u}_1^b}{\bar{u}_2^b} = \frac{\bar{p}^s}{\bar{p}^b} = \frac{1 - \bar{b}^b}{\bar{s}^b} \tag{11}$$

This, of course, is the familiar fact that the ratio of the marginal utilities from steel and bread must equal the ratio of the respective prices at the maximum. Furthermore, we require that θ be such that the utilities in the two cities are equal to each other, i.e., $u(s^s, b^s, \theta) = u(s^b, b^b, 0)$. If this were not the case, people would move from one city to another. This would violate our notion of equilibrium since total welfare could be increased by such migration. We denote the solution values of our equilibrium conditions by $\bar{s}^s$, $\bar{b}^s$, $\bar{\theta}$, $\bar{s}^b$, and $\bar{b}^b$.

III. Conditions for a Pareto Optimum

Pareto optimality will be achieved if each consumer's utility is a maximum given the utility levels of all other consumers. Following the approach set forth by Strotz, we shall alternatively interpret the problem so as to maximize the linear social welfare function,

$$\sum_s u(s^s, b^s, \theta) + \sum_b u(s^b, b^b, \theta) = \theta u(s^s, b^s, \theta) + (1 - \theta) u(s^b, b^b, 0)$$

given the available resources (total labor supply) and the production alternatives open to society (equations (3) and (4)). The equality comes from the fact that all individuals are identical, i.e., the welfare weights are assumed to be unity. Mathematically we set up the Lagrangean:

$$\begin{aligned}\theta u(s^s, b^s, \theta) &+ (1 - \theta) u(s^b, b^b, 0) \\ &+ \mu[\theta b^s + (1 - \theta) b^b - (1 - \theta)] \\ &+ \gamma[\theta s^s + (1 - \theta) s^b - \theta] \\ &+ \delta[u(s^s, b^s, \theta) - u(s^b, b^b, 0)]\end{aligned} \tag{12}$$

where μ, γ, and δ are Lagrange multipliers insuring that the constraints are not violated. Note the last constraint is derived from our requirement that the utilities in the two cities be equal to each other.

Differentiating (12) with respect to s^s, b^s, s^b, b^b, and θ, respectively:

$$u_1^s(\theta + \delta) + \gamma\theta = 0 \tag{13}$$

$$u_2^s(\theta + \delta) + \mu\theta = 0 \tag{14}$$

$$u_1^b(\theta - \delta) + \gamma(1 - \theta) = 0 \tag{15}$$

$$u_2^b(\theta - \delta) + \mu(1 - \theta) = 0 \tag{16}$$

$$\begin{aligned}u(s^s, b^s, \theta) - u(s^b, b^b, 0) \\ + \theta u_3^s + \mu(b^s - b^b + 1) \\ + \gamma(s^s - s^b - 1) + \delta u_3^s = 0\end{aligned} \tag{17}$$

Dividing (13) by (14) and (15) by (16) we obtain:

$$\frac{u_1^s}{u_2^s} = \frac{u_1^b}{u_2^b} = \frac{\gamma}{\mu} \tag{18}$$

Rearranging (17), using (14), the production constraints (equations (3) and (4)), and the requirement that the utilities in the two cities are equal, we find:[12]

$$\frac{\gamma}{\mu} = \frac{1 - b^b}{s^b} = \frac{u_3^s \theta^2}{u_2^s s^b} \tag{19}$$

Equating (18) and (19) we derive a necessary condition for an optimum.[13]

[11] The last equality is derived from the income constraint for bread city residents in equation (6). The partial derivatives of the steel city and bread city utility functions are evaluated at ($\bar{s}^s$, $\bar{b}^s$, $\bar{\theta}$) and ($\bar{s}^b$, $\bar{b}^b$), respectively.

[12] In this expression, s^b has the units of steel, not steel per person. See fn. 3.

[13] This is the price ratio which will induce "optimal" behavior from the consumers in the two cities. A hat over a variable denotes the value attained at the optimum. In this case, the derivatives of the steel-city and bread-city utility functions are evaluated at ($\hat{s}^s$, $\hat{b}^s$, $\hat{\theta}$) and ($\hat{s}^b$, $\hat{b}^b$, 0), respectively.

$$\frac{\hat{u}_1^s}{\hat{u}_2^s} = \frac{\hat{u}_1^b}{\hat{u}_2^b} = \frac{1 - \hat{b}^b}{\hat{s}^b} - \frac{\hat{u}_3^s \hat{\theta}^2}{\hat{u}_2^s \hat{s}^b} = \frac{\hat{p}^s}{\hat{p}^b} \tag{20}$$

IV. Individual Maximization and Pareto Optimality

We now ask whether individual utility maximization, subject to fixed unit prices common to all individuals, will lead to a Pareto optimum.[14] To do this it is sufficient to compare $\bar{\theta}$ with $\hat{\theta}$. If they are not equal, production will be incorrect, i.e., too much or too little steel will be produced. More precisely, we shall prove that $\bar{\theta} > \hat{\theta}$, i.e., optimality requires a redistribution of population from steel city to bread city with less steel production (and less pollution).

First, we examine the expenditures of a bread city resident in the two situations in terms of bread units:

$$E_b^b = b^b + \frac{p^s}{p^b} s^b \tag{21}$$

where E_b^b represents the number of units of bread a resident could buy if no steel was bought.

Substituting the price ratio corresponding to individual maximization (using equation (11)) into equation (21) we obtain:

$$\bar{E}_b^b = 1 \tag{22}$$

Hence, the income of bread workers (in this case, the price of bread) is equal to unity.

Similarly, we use the price ratio at an optimum (from equation (20)) and substitute it into equation (21) and find:

$$\hat{E}_b^b = 1 - \frac{\hat{u}_3^s \hat{\theta}^2}{\hat{u}_2^s} \tag{23}$$

Remembering that $u_3 < 0$, we see optimality requires that the income in units of bread for bread workers exceed unity.[15]

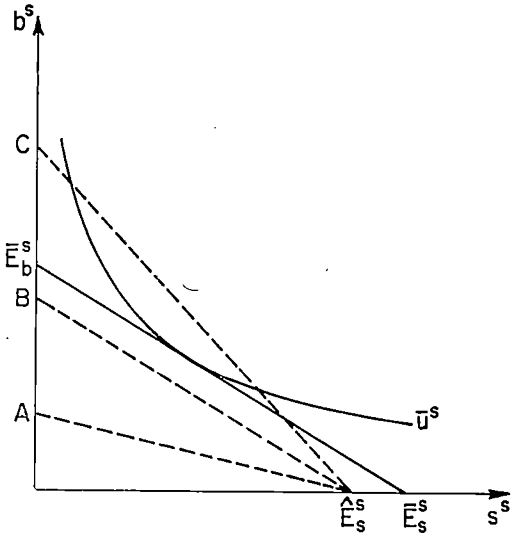


FIGURE 1

Figure 1 illustrates the situation for a representative individual of steel city. The axes are labeled his consumption of bread and steel and an indifference curve is drawn at the equilibrium point under individual maximization tangent to the equilibrium price ratio.[16] We have just stated (fn. 15) that under optimality the income of steel city residents in terms of steel units falls, hence, we can arbitrarily pick a point to the left of $\bar{E}_s^s$ and label it $\hat{E}_s^s$. We then graph all three possibilities for the change in relative prices from individual maximization to optimality: the price ratio can fall, $(A\hat{E}_s^s)$, remain constant, $(B\hat{E}_s^s)$, or increase, $(C\hat{E}_s^s)$. We know that the level of utility which the individual attains under optimality must be at least as high as his utility under individual maximization; hence if the price ratio either falls or remains constant, pollution (and steel production) must decrease.[17]

Thus, we will limit ourselves to the cases in which the price ratio might increase. From

[14] Note that the Pareto situation is a generalization of the case of individual utility maximization with the removal of the constraint that expenditures equal wage income. Therefore, by the Le Chatelier principle, the Pareto situation must be at least as good as individual maximization; if it differs it must be superior.

[15] Using the fact that $(1-b^b)/s^b = b^s/(1-s^s)$ and that $(1-b^b)/s^b - u_3^s\theta^2/u_2^s s^b = b^s/(1-s^s) - u_3^s\theta(1-\theta)/u_2^s(1-s^s)$ we can likewise show that $\bar{E}_s^s > \hat{E}_s^s$.

[16] The figure is only in the dimensions of bread and steel; hence, each indifference curve has a level of pollution, θ, associated with it.

[17] Mathematically, if $\hat{p}^s/\hat{p}^b \leq \bar{p}^s/\bar{p}^b$ then $\bar{\theta} > \hat{\theta}$ in order that $\hat{u}^s \geq \bar{u}^s$.

equations (11) and (20) this means we will be examining cases for which the following holds:

$$(24)\qquad \frac{\bar{u}_1^s}{\bar{u}_2^s} = \frac{\bar{u}_1^b}{\bar{u}_2^b} < \frac{\hat{u}_1^s}{\hat{u}_2^s} = \frac{\hat{u}_1^b}{\hat{u}_2^b}$$

Using (24), the production constraints (equations (3) and (4)), the fact that $\hat{u}^s = \hat{u}^s \geq \bar{u}^s = \bar{u}^b$, the assumption of diminishing marginal utility, and the neutrality of pollution,[18] we can enumerate the admissible changes in our variables:[19]

s^b	+	+	+	−	−	−	−	−	0	0	0
b^b	+	+	+	+	+	+	+	+	+	+	+
θ	−	−	−	−	−	−	−	−	−	−	−
s^s	−	−	−	0	−	−	−	+	−	−	−
b^s	0	+	−	+	0	+	−	+	0	+	−

We see the only possible changes involve a reduction in θ. Without some externally imposed adjustment, i.e., tax, the results from individual utility maximization will not adequately reflect the externality associated with steel production. Individuals moving to steel city will not take account of the extra pollution they subject all steel city residents to and the externality will not be completely internalized. Thus, there will be a misallocation of resources (labor) between the two cities.

[18] By neutrality, we mean that a change in the pollution level does not affect the relative marginal utility of bread and steel consumption.

[19] The table indicates the direction of change of the variable from individual maximization to optimality, e.g., s^b followed by a + means that $\hat{s}^b > \bar{s}^b$. For example, looking at the first column, we see that both s^b and b^b may increase so as to satisfy the relation in (24) pertaining to bread city, i.e., that the ratio of marginal utilities increases. At the same time, it is clear that this will mean that $\hat{u}^b \geq \bar{u}^b$ since more of both goods are being consumed. A consistent solution for steel city would involve a reduction in s^s with b^s remaining constant. This would necessitate a decrease in θ. θ could not increase or remain unchanged if both s^b and b^b rise and our relationships are satisfied. Similar arguments can be made for the other columns in the table.

V. The Tax

Comparing equations (22) and (23), we see that an income transfer from residents of steel city to residents of bread city must take place to attain optimality from individual maximization. The magnitude of this transfer (in terms of bread units) is given by:

$$(25)\qquad \tau = \hat{E}_b^b - \bar{E}_b^b = -\frac{\hat{u}_3^s \hat{\theta}^2}{\hat{u}_2^s} = -\frac{\hat{u}_3^s}{\hat{u}_2^s/\hat{\theta}}\,\hat{\theta} > 0$$

This quantity can be viewed as a tax on and subsidy to the factors of production.[20] Thus, steel producers are taxed for employing the "pollution-creating" input (steel labor), while bread producers are subsidized for utilizing the "pollution-free" input (bread labor). The resulting new demand curves for labor lead to an optimal reallocation of resources.

In our case, there is another interpretation to the assessment. The numerator of the tax collected from everyone in steel city (θ people) represents the marginal disutility per person (for all people) of a change in the pollution level and the denominator represents the marginal utility of bread per person in steel city. Hence, the ratio measures the disutility of pollution relative to the utility of bread for people in steel city. When multiplied by the population of steel city, this reflects the total amount of bread they should be willing to give up to decrease the level of pollution.

VI. A Numerical Example

This conclusion can be made clearer through the use of a numerical example. We shall derive the results for individual utility maximization and for a Pareto optimum utilizing the following Cobb-Douglas utility functions:[21]

$$(26)\qquad u^s = \bar{e}^{\theta}(s^s)^{1-\beta}(b^s)^{\beta} \quad \text{(steel city)}$$

$$0 < \beta < 1$$

$$(27)\qquad u^b = (s^b)^{1-\beta}(b^b)^{\beta} \quad \text{(bread city)}$$

[20] Alternatively, it can be looked upon as an inducement to the labor force to change their working patterns. See fn. 4.

[21] Note, for bread city, $\theta = 0$ and $\bar{e}^{\theta} = 1$ which results in (27).

We can now use these functions, together with the production constraints (equations (3) and (4)), the budget constraints (equations (5) and (6)), and the condition in equation (11) to solve a system of simultaneous equations for the unknowns $\bar{s}^s$, $\bar{s}^b$, $\bar{b}^s$, $\bar{b}^b$, and $\bar{P}$ in terms of β and θ:[22]

$$\bar{s}^s = 1 - \beta \tag{28}$$

$$\bar{s}^b = \frac{\beta\bar{\theta}}{1-\bar{\theta}} \tag{29}$$

$$\bar{b}^s = \frac{(1-\bar{\theta})(1-\beta)}{\bar{\theta}} \tag{30}$$

$$\bar{b}^b = \beta \tag{31}$$

$$\bar{P} = \frac{(1-\beta)(1-\bar{\theta})}{\beta\bar{\theta}} \tag{32}$$

Substituting these values into (26) and (27) and remembering that in equilibrium the utilities in the two cities must be equal we obtain:

$$e^{-\bar{\theta}}\left(\frac{1-\bar{\theta}}{\bar{\theta}}\right) = \frac{\beta}{1-\beta} \tag{33}$$

Letting $\beta=.38$, we can solve (28) through (33) for approximate numerical values of the variables. We find that steel consumption per capita in steel city ($\bar{s}^s$) is .62, steel consumption per capita in bread city ($\bar{s}^b$) is .38, bread consumption per capita in steel city ($\bar{b}^s$) is .62, and bread consumption per capita in bread city ($\bar{b}^b$) is .38. The relative price ratio ($\bar{P}$) is 1.63. The total population is equally distributed between the two cities ($\bar{\theta}=.50$).[23] Finally, the level of utility for each individual in society ($\bar{u}^s=\bar{u}^b$) is .38.

Now, we solve for the optimum values using the Cobb-Douglas utility functions and relation (20):

$$\hat{s}^s = \frac{1-\beta}{1+\hat{\theta}-\hat{\theta}^2} \tag{34}$$

$$\hat{s}^b = \frac{\hat{\theta}}{1-\hat{\theta}}\left(\frac{\beta+\hat{\theta}-\hat{\theta}^2}{1+\hat{\theta}-\hat{\theta}^2}\right) \tag{35}$$

$$\hat{b}^s = \frac{1-\hat{\theta}}{\hat{\theta}}\left(\frac{1-\beta}{1+\hat{\theta}-\hat{\theta}^2}\right) \tag{36}$$

$$\hat{b}^b = \frac{\beta+\hat{\theta}-\hat{\theta}^2}{1+\hat{\theta}-\hat{\theta}^2} \tag{37}$$

$$\hat{P} = \frac{(1-\beta)(1-\hat{\theta})}{\beta\hat{\theta}} \tag{38}$$

Again we substitute into the utility functions (equations (26) and (27)) and obtain:

$$e^{-\hat{\theta}}\left(\frac{1-\hat{\theta}}{\hat{\theta}}\right) = \frac{\beta}{1-\beta} + \frac{\hat{\theta}-\hat{\theta}^2}{1-\beta} \tag{39}$$

For $\beta=.38$, we solve (34) through (39) for the numerical values of the unknowns. In this case, steel consumption per capita in steel city ($\hat{s}^s$) is .50, steel consumption per capita in bread city ($\hat{s}^b$) is .33, bread consumption per capita in steel city ($\hat{b}^s$) is .75, and bread consumption per capita in bread city ($\hat{b}^b$) is .50. The relative price ratio ($\hat{P}$) is 2.45. Less of the total population now resides in steel city ($\hat{\theta}=.40$); hence, there is less steel production and less pollution. The level of utility for each individual ($\hat{u}^s=\hat{u}^b$) has now increased to .39. The tax (τ) imposed on steel city residents and transferred to each bread city inhabitant is .32.

VII. Discussion

We have posed a simple market situation in which an externality (air pollution) arises as an explicit joint product in the manufacture of a "normal" good. People engaged in the production of the good are subjected to the externality, but are free to move to another pollution-free environment and produce a different good.[24] We have shown that

[22] Since we are only concerned with relative prices, we define $P=p^s/p^b$. Note, too, that initially we are specifying a given distribution of population θ (and; hence, income). We will then search over θ so as to equate the individual utilities in the two cities.

[23] This, of course, is also the level of pollultion in steel city.

[24] Our model has dealt with the polar case of completely separate production locations. In this way, we allowed the market maximum freedom in coping with the problem; it was unable to do so in an optimal manner. Permitting bread production to take place in steel city would further aggravate the situation.

the market fails to allocate optimally the resources (population) without outside intervention. In particular, a tax on steel city residents or a subsidy to bread city residents is needed. If this activity is to be neutral with respect to distributional effects, there must be both a tax on steel city residents *and* a subsidy to bread city residents.

The essential element in the model which causes the market to fail is the fact that the externality takes the form of a pure public good (or, more precisely, a "pure public bad"). Allowing some people the choice of consuming more goods and subjecting themselves to more pollution does not lead to a Pareto optimum. The final person moving to steel city does not take account of the additional pollution he inflicts upon *all* steel city residents.

This result is of interest when one considers suggestions that we allow some cities to become the sewers for American production or that we export our pollution by importing pollution-producing goods. Clearly, these solutions are nonoptimal for society as a whole. Apparently, sewer city would be too polluted for Pareto optimality.

APPENDIX[25]

The nonoptimality theorem was proved under the rather strong assumptions that all individuals possessed identical cardinal utility functions. For it will be recalled that the original Pareto optimality conditions were derived by summing the individual utility functions. Here we allow each individual to possess his own differentiable utility function defined only up to a smooth monotonic transformation. Again, we assume positive marginal utilities for bread and steel, and negative marginal utilities for pollution. Moreover, we will use only the signs, not the marginal utilities in our derivation, and these signs are invariant under monotonic transformations of the functions. Hence, our results do not depend upon the cardinal scales of the utility functions, and we can regard these functions as defined ordinally.

Let each individual, i, have a differentiable utility function:

$$u^i = u^i(s^i, b^i, a^i) \tag{A1}$$

Again,

(A2) $a^i = 0$ for residents of bread city ($i \in B$)

(A3) $a^i = \theta$ for residents of steel city ($i \in S$)

Each person is constrained to spend no more than his income, which is equal to the value of his product plus (minus) his subsidy (assessment):[26]

$$p^i - p^s s^i - b^i + e^i = 0 \tag{A4}$$

where $p^i = 1$ for $i \in B$; $p^i = p^s$ for $i \in S$; and e^i is the subsidy (assessment) of the ith individual.[27]

Necessary conditions for competitive equilibrium again imply:

$$\frac{u_1^i}{u_2^i} = p^s \qquad \text{for all } i \tag{A5}$$

Now assume that some individuals are moved from steel city to bread city, so that there is a decrease, $d\theta$, in the population of the former city and in the production of steel, and a corresponding increase in the population of the latter city and in the production of bread. As a result of the decrease in steel production there is an increase, dp^s, in the relative price of steel.[28] We now calculate the consequent changes in utility:

$$du^i = u_1^i ds^i + u_2^i db^i + u_3^i da^i \tag{A6}$$

where $da^i = 0$ for $i \in B$, and $da^i = d\theta$ for $i \in S$.

From (A4), the derivative of the budget constraint with no subsidy or assessment is:

$$dp^i - p^s ds^i - s^i dp^s - db^i = 0 \tag{A7}$$

[25] Herbert A. Simon is coauthor of the Appendix.

[26] We assume that the government may assess (subsidize) any member of the population, and that the assessments (subsidies) may be different for different persons.

[27] Note, for simplification, we take the price of bread as numeraire, while the price of steel (per unit of bread) is p_s.

[28] We assume that a decrease in the production of steel and an increase in the production of bread always causes an increase in the price of steel relative to the price of bread. This condition is stated independently of the utility functions.

Using (A5) and (A7), we can rewrite (A6) as:

$$\text{(A8)} \quad \frac{du^i}{u_2^i} = dp^i - s^i dp^s + \left(\frac{u_3^i}{u_2^i}\right) da^i$$

or, more explicitly, for B and S separately:

$$\text{(A9)} \quad \frac{du^i}{u_2^i} = -s^i dp^s \qquad (i \in B)$$

$$\text{(A10)} \quad \frac{du^i}{u_2^i} = (1-s^i)dp^s + \left(\frac{u_3^i}{u_2^i}\right) d\theta \qquad (i \in S)$$

Since $dp^s > 0$ and $s^i > 0$, we see from (A9) that the population shift has lowered the utility of all residents of bread city. Since $s^i < 1$, $u_3^i < 0$, $u_2^i > 0$ and $d\theta < 0$, we see from (A10) that the population shift has increased the utility of all residents of steel city. Moreover, since the dimensionality of the terms in (A9) and (A10) is "loaves," we can determine immediately how large a subsidy or assessment, in the same units, would restore utilities to their previous levels.

We now give a subsidy to or impose an assessment on each person of such magnitude that his utility will be at least equal to its initial level.[29] If this scheme is to be feasible then the sum of the assessments must be equal to or greater than the sum of the subsidies (the government cannot redistribute more than it collects).

We wish to show that:

$$\text{(A11)} \quad \sum_U e^i = \sum_B e^i + \sum_S e^i < 0$$

[29] The quantities, e^i, are simply calculated by replacing the term on the left of (A8) (or equivalently (A9) and (A10)) by $-e^i$ (the minus sign so that the subsidy or assessment will be of opposite sign to the change in utility).

Using (A9) and (A10), with the proper substitutions (fn. 29), and reversing signs on both sides of the equations, we obtain:

$$\text{(A12)} \quad \sum_B e^i = dp^s \sum_B s^i$$

$$\text{(A13)} \quad \sum_S e^i = -\theta dp^s + dp^s \sum_S s^i - d\theta \sum_S \frac{u_3^i}{u_2^i}$$

(recalling that S contains θ persons).

Now, summing (A12) and (A13):

$$\text{(A14)} \quad \sum_U e^i = -d\theta \sum_S \frac{u_3^i}{u_2^i} < 0$$

(recalling that $\sum_B s^i + \sum_S s^i = \sum_U s^i = \theta$).

Thus, with this system of subsidies and assessments, which leaves all utilities unchanged from their initial levels, the government is left with a surplus which it can distribute among all persons so that all utilities will be higher than in the initial state.[30]

REFERENCES

W. J. Baumol, "On Taxation and the Control of Externalities," *Amer. Econ. Rev.*, June 1972, *62*, 307–22.

R. H. Coase, "The Problem of Social Cost," *J. Law Econ.*, Oct. 1960, *3*, 1–44.

A. C. Pigou, *The Economics of Welfare*, 4th ed., London 1932.

R. Strotz, "Urban Transportation Parables," in J. Margolis, ed., *The Public Economy of Urban Communities*, Washington 1965.

[30] Note, the surplus is exactly equal to the market value (i.e., bread equivalent) of the benefits obtained by residents of steel city from the lower pollution resulting from the reduction in population and steel production. The higher costs of steel to residents of bread city can be exactly compensated by assessing the higher cash incomes of residents of steel city.

Expectations and the Labor Supply

By CHARLES L. HEDRICK*

The long-run and short-run dynamics of the supply of national output are a key to macro-economic behavior. These dynamics form the basis of the "natural rate" analysis of Milton Friedman and others. And the key to the dynamics of the total supply is that of the labor supply. It would be an exaggeration to claim that capital supply and the behavior of firms is irrelevant to macroeconomics. But output can be viewed as the product of a cyclical employment variable with a more or less uniformly increasing productivity. Thus it is clear that the properties of the total supply are highly dependent upon those of the labor supply. The purpose of this paper is to explore the dynamics of the labor supply, with particular emphasis on the effect of expectations models. In the process a regression that Robert Lucas and Leonard Rapping used to explain the labor supply is replicated with more recent data. Although further analysis is done, one of the purposes of this study was to verify that their earlier 1930–65 results also hold for a more recent period (1950–70), and with improved data series.

I. The Labor Supply Equation

The basic labor supply equation to be used was put forward by Lucas and Rapping as part of a dynamic model of employment and unemployment. It was derived by assuming that households maximize utility by trading off present and future consumption and leisure. After aggregating over the entire economy and omitting the variables they found not to add significant explanatory power in the regressions, the following log-linear approximation resulted:

$$(1)\qquad N_t - M_t = b_0 + b_1 w_t - b_2 w_t^* + b_3(P_t - P_t^*)$$

Here, the b are positive constants, N represents the total employment in man-hours, M the population adjusted for age-sex distribution, w the real wage rate and P the price level. As is true throughout this paper, the variables are logarithms of the quantities represented.

The starred variables are actually intended to represent estimates of all future labor and consumption possibilities considered by the household. However, for computational convenience, and because no good model is at hand for making long-term predictions, in practice one may use a one-period planning horizon, giving

$$(2)\qquad w_t^* = E(w_{t+1}\mid t)$$

and a similar relation for P. The right-hand side of this equation is supposed to represent the expectation of w_{t+1}, given the values of the variables up to time t. This interpretation of the starred variables appears to match the intention of Lucas and Rapping's derivation. An alternative is available, however:

$$(3)\qquad w_t^* = E(w_t\mid t-1)$$

This interpretation is more consistent with the "natural rate hypothesis," and is also mentioned by Lucas and Rapping. Note that there is considerable conceptual difference between (2) and (3). Since b_1 and b_2 turn out to be nearly equal, using equation (2) would imply that employment is determined by $w_t - E(w_{t+1}\mid t)$, the negative of the expected increase in w between t and t+1. On the other hand, equation (3) would give $w_t - E(w_t\mid t-1)$, the degree to which the pres-

* Carnegie-Mellon University. I would like to acknowledge the helpful comments of Robert Lucas, Edward Prescott, and Leonard Rapping, all of Carnegie-Mellon University, and referees for this journal, on earlier drafts of this paper. Particular thanks are due to Prescott for the use of his adaptive regression program. I was supported during this research by a National Science Foundation Graduate Fellowship.

ent period differs from expectation ("surprise").[1] Although the regressions done by Lucas and Rapping could not distinguish these two interpretations, it will be seen later that it is possible to distinguish them. All that remains before an equation may be written which is sufficiently explicit for empirical test is to specify a model for expectations formation. The model used by Lucas and Rapping is adaptive expectations,[2]

$$(4)\quad E(w_t \mid t-1) = \lambda w_{t-1} + (1-\lambda)E(w_{t-1} \mid t-2) + \lambda'$$

and similarly for P. This is intended primarily as a psychological model, the expectation being read as actual subjective expectations, not any objective statistical estimator.[3] In the context of the natural rate

[1] Of course Lucas and Rapping's derivation of equation (1) no longer applies with this second interpretation. As will become clear later, the crucial difference between $E(w_t|t-1)$ and $E(w_{t+1}|t)$ is not the period predicted, but the period on which the prediction is based. Thus one might consider using $E(w_{t+1}|t-1)$. Because it is based on information from $t-1$ it will have the properties of the second interpretation and be empirically indistinguishable from it. But since it predicts w_{t+1}, it can be used with the original derivation. Another possibility for motivating equation (1) under the second interpretation is to begin directly with the assumption that there is a natural rate of unemployment. (See Donald Gordon and Allan Hynes for arguments relating to this.) When real wages or prices rise unexpectedly, workers will be deceived about what acceptance wage rate will maintain their optimum period of unemployment and will accept jobs more quickly than normal, hence raising employment. This is similar to the sort of job search model used, for example, by Dale Mortensen. However, Gordon and Hynes' arguments do not seem to be based on job search per se. Their view would allow some period of unemployment to be an optimal strategy even with perfect information if job offers at different wages become available randomly.

[2] Note that with this model, since w_t is not known at $t-1$,

$$E(w_{t+1} \mid t-1) = \lambda w_{t-1} + (1-\lambda)E(w_{t-1} \mid t-2) + 2\lambda'$$

Since this differs from $E(w_t|t-1)$ only by a constant it will be indistinguishable from it in the regressions. This justifies the comment made in fn. 1 about the equivalence of $E(w_{t+1}|t-1)$ and $E(w_t|t-1)$.

[3] John Muth showed that equation (4) is in fact the statistically optimal way to estimate w_t in the case that a natural rate mechanism is really applicable. This case may be expressed by

[3] Continued:

$$w_t = x_t + \lambda' + u_t$$
$$x_t = x_{t-1} + v_t$$

where x_t is the natural rate, and u_t and v_t are each a series of identically distributed serially independent random variables with mean zero. Although u_t and u_{t-1} (and also v_t and v_{t-1}) must be independent, u_t and v_t may be correlated to any degree, even perfectly. As is stated above, it is not necessary that these equations actually hold, since equation (4) is intended only as a model of subjective expectations. However, it seemed of some interest to know how well equation (4) would predict the actual wages and prices. Thus a non-linear regression was done on (4) in the form

$$\text{(a)}\quad w_t = \lambda w_{t-1} + (1-\lambda)w^*_{t-1} + \lambda' + u_t$$
$$\text{(b)}\quad w^*_t = \lambda w_{t-1} + (1-\lambda)w^*_{t-1} + \lambda'$$

λ and λ' were adjusted until the square error in (a) was minimized, (b) being used to define w^*_t. Several years before the beginning of the sample w^*_t was set equal to the actual value of w_t and equation (b) was iterated until the first value of w^*_t needed was found. The results of this non-linear regression show λ (the speed of adjustment) to be .60 for wages and 1.78 for prices. The trend terms, λ', are .031 and .026, respectively. This was based on the data referred to elsewhere in this report as *BLS* (see the Appendix for details). Satisfyingly large r-squares (.996 and .989) and small errors (.17 and .05 for λ, .002 and .005 for λ') were obtained, and the results appear insensitive to the initial approximation. But trials over subperiods of the 1950–70 range show that the uncertainties are considerably larger than the standard error estimates. For example using 1952–70 gives a λ_w of .92, and 1951–70 gives a λ_P of 2.20. Both of these are far outside the error limits of the 1950–70 results. The fact that λ_P is greater than 1.00 suggests that no natural rate mechanism is at work there. Rather than a tendency for prices to return to normal, apparently the reverse tendency exists! It is interesting in this context to note the results of the following three-equation macro-economic model:

$$\text{(a}'\text{)}\quad y_t = a(P_t - E(P_t \mid t-1))$$
$$\text{(b}'\text{)}\quad y_t + P_t = x_t$$
$$\text{(c}'\text{)}\quad x_t = x_{t-1} + u_t$$

where y_t is real national output, x_t is demand for nominal national output, and u_t is a series of identically distributed serially independent random variables with mean zero. Equation (a′) is just equation (1) ignoring the real wage term; (b′) asserts equality of supply and demand for national output; and (c′) represents demand for output as being a random walk, trends having been suppressed. The solution to these three equations can be expressed recursively as

$$E(P_t \mid t-1) = (1+a)P_{t-1} - aE(P_{t-1} \mid t-2)$$

Note that this has the form of adaptive expectations, but λ is greater than 1.00. (See Lucas for a full development of a similar model.)

hypothesis the expectation is to be seen as a slowly changing natural rate. λ represents the degree to which the predictions weight the actual past performance rather than the previous prediction (or natural rate).

II. Testing the Adaptive Expectations Model

The equation for adaptive expectations, (4), can be combined with (1) by a Koyck transformation to give a single equation suitable for estimation. The type of transformation used assumes that λ has the same value for w and P. Using the interpretation of w_t^* as $E(w_{t+1}|\,t)$, and the equivalent for P, gives

$$(5)\quad \begin{aligned} N_t - M_t = {} & (b_0\lambda - b_2\lambda' - b_3\lambda'') \\ & + (b_1 - \lambda b_2)w_t \\ & - (1-\lambda)b_1 w_{t-1} \\ & + (1-\lambda)b_3(P_t - P_{t-1} \\ & + (1-\lambda)(N_{t-1} - M_{t-1}) \end{aligned}$$

Using $E(w_t|\,t-1)$ gives

$$(6)\quad \begin{aligned} N_t - M_t = {} & (b_0\lambda - b_2\lambda' - b_3\lambda'') + b_1 w_t \\ & - (b_2\lambda + b_1(1-\lambda))w_{t-1} \\ & + b_3(P_t - P_{t-1})) \\ & + (1-\lambda)(N_{t-1} - M_{t-1}) \end{aligned}$$

Note that if all one has is regression results on these equations, there will be no way to tell equations (5) and (6) apart. The decision between the two interpretations of expectation will have to be based on other evidence.

Table 1 presents results of regressions based on these two equations, for three different sets of data. The first is Lucas and Rapping's original 1930–65 data, the results being taken directly from their paper. The second is their data extended by the author to cover the 1950–70 period. The third set of data covers the same 1950–70 period, but is based on data from the Bureau of Labor Statistics (*BLS*) not available for earlier periods. (See the Appendix for more details on the data.)[4] Knowledge of the way the data was derived, as well as the results in Table 1, suggest that the *BLS* data is more accurate, or at least more sensitive to short-term variations. This would explain the larger values of the three coefficients, the smaller values of the coefficient of the lagged dependent variable (which may have more weight thrown on it than that due to $(1-\lambda)$ when the model is bad), and the higher r-square in the estimation using the *BLS*

[4] Here, as in all other regression results, asymptotic estimates of the standard errors are given underneath each coefficient in parentheses. Also the r-square and Durbin-Watson statistics are given. In fact it is not certain just what the standard errors mean in simultaneous equation models with such a small sample size. However, I will follow the usual practice of ignoring this problem and treat the standard errors as real estimates of the standard errors of presumed normal distributions of the coefficients. Since there is little alternative, other than giving up any attempt to determine whether coefficients of two different regressions differ significantly, one may only hope that this interpretation is not misleading. Note that the usefulness of the estimates is even more dubious for the non-linear results described in fn. 3 and below.

TABLE 1—TWO-STAGE LINEAR REGRESSION RESULTS FOR KOYCK-TRANSFORMED LABOR SUPPLY EQUATION[a]

	Regression Coefficient of $N_t - M_t$ on:					
Data	w_t	w_{t-1}	$P_t - P_{t-1}$	$(N-M)_{t-1}$	r^2	D-W
Lucas-Rapping 1930–65	1.40 (.51)	−1.39 (.51)	.74 (.17)	.64 (.09)	.80	1.56
Lucas-Rapping, extended by author 1950–70	1.44 (.52)	−1.43 (.50)	.99 (.30)	.33 (.16)	.62	1.91
BLS 1950–70	1.99 (.24)	−2.18 (.24)	1.05 (.14)	.20 (.09)	.98	1.93

[a] Standard errors are shown in parentheses.

TABLE 2—LINEAR REGRESSIONS FOR EVALUATING EFFECT OF TWO-STAGE LEAST SQUARES[a]

	Regression Coefficients (−1 for dependent variable):										
	N_t-M_t	w_t	w_{t-1}	(P_t-P_{t-1})	$(N-M)_{t-1}$	$(N+Q-y)_{t-1}$	(y_t-y_{t-1})	Q_t	y_t-M_t	r^2	D-W
Reduced-form wage equation, Lucas-Rapping data, 1950–70		−1	.73 (.15)	−.32 (.10)	−.24 (.45)	.56 (.42)	−.37 (.44)	.14 (.64)	.51 (.44)	.998	2.06
Reduced-form wage equation, *BLS* data, 1950–70		−1	.85 (.22)	−.37 (.15)	.39 (.44)	−.18 (.43)	.30 (.46)	.42 (.44)	.01 (.48)	.998	2.61
Koyck-transformed labor supply equation, ordinary least squares, Lucas-Rapping data, 1950–70	−1	.82 (.48)	−.83 (.46)	.83 (.33)	.40 (.17)					.518	1.89
Koyck-transformed labor supply equation, ordinary least squares, *BLS* data, 1950–70	−1	.91 (.32)	−1.12 (.32)	.69 (.24)	.23 (.18)					.940	2.16

[a] Standard errors are shown in parentheses.

data. For this reason the *BLS* data is used elsewhere in the paper. Lucas and Rapping's data was extended to 1970 to allow comparison between the two periods using the same data sources. It can be seen that the coefficients of wages, prices and expectations are nearly the same for the two periods, but that of the lagged dependent variable is not.

Table 1 shows two-stage least square regressions. The corresponding ordinary least square results are given in Table 2 for comparison, along with the reduced-form wage equation used for the wage term in the two-stage process. Since both wages and employment are determined in the labor market, there is undoubtably correlation between the wage and error terms in the ordinary least squares. The extent of this effect may be seen by comparing the two-stage results in Table 1 with the ordinary results in Table 2. It can be seen that the coefficients are larger (and more significant) and the r-square is larger for the two-stage.

It was suggested above that when the equation (5) or (6) does not fit the data very well, more weight is thrown on the coefficient of the lagged dependent variable than just $(1-\lambda)$. If so, it would be useful to have a way of testing equation (1) that did not introduce the lagged dependent variable. In fact such a way exists. The second line in Table 3 gives the results of a non-linear regression based on equation (1) directly, with w_t^* and P_t^* given by

$$(7)\quad E(w_t \mid t-1) = \lambda w_{t-1} + (1-\lambda)E(w_{t-1} \mid t-2)$$

and similarly for P.

The computer program used to produce those results searched for the values of λ_w and λ_P that minimize the squared error in equation (1) when equation (7) is used to define the expectations.[5] The agreement between these results and those based on equations (5) and (6) (repeated as the first line of Table 3) is encouraging. Note that the coefficient of the lagged dependent variable represents $(1-\lambda)$, hence λ is inferred to be .80, between the values of λ_w and λ_P coming from the non-linear regression.[6]

[5] Equation (7) differs from the equations (a), (b) in fn. 3 mainly in the lack of the trend term. This is incorporated in the intercept of the regression on equation (1), and is not separately identifiable. As in fn. 3, w_t^* is set to the actual value of w_t in some year before the start of the regression. Repeat runs using different years show that the results are not sensitive to what beginning period is chosen.

[6] These results have been given with the interpretation of P_t^* as $E(P_t \mid t-1)$, which is implicit in equation (7). The equivalent for $E(P_{t+1} \mid t)$ is

$$E(P_{t+1} \mid t) = P_t + (1-\lambda)E(P_t \mid t-1)$$

From this we see that $P_t - P_t^*$ would become

TABLE 3—TWO-STAGE REGRESSIONS ON VARIOUS VERSIONS OF LABOR SUPPLY EQUATIONS, *BLS* DATA[a]

	Regression Coefficient of N_t-M_t on:							
	w_t	w_{t-1}	P_t-P_{t-1}	$(N-M)_{t-1}$	λ_w	λ_P	r^2	*D-W*
Koyck-transformed equation, linear regression, 1950–70	1.99 (.24)	−2.18 (.24)	1.05 (.14)	.20 (.09)			.98	1.93
Non-linear regression on λ's linear on rest, 1950–70	2.13 (.23)	−2.37 (.22)	1.11 (.13)		.95 (.07)	.76 (.10)	.99	1.49
Equation with λ = 1.00, linear regression, 1950–70	1.55 (.25)	−2.10 (.29)	.84 (.14)				.97	1.39
Koyck-transformed equation, linear regression, 1952–70	2.06 (.20)	−2.35 (.20)	1.60 (.23)	−.05 (.11)			.99	1.82
Non-linear regression on λ's, linear on rest, 1952–70	2.07 (.20)	−2.34 (.20)	1.44 (.19)		.99 (.07)	.93 (.10)	.99	1.91
Equation with λ = 1.00, linear regression, 1952–70	1.68 (.20)	−2.27 (.24)	1.02 (.17)				.98	2.07

[a] Standard errors are shown in parentheses.

There is a reservation about these results, however. Similar regressions were done for subperiods within the 1950–70 time span. They suggest that the error estimates for the λ may be understating the uncertainty. The values of λ from the linear regression are distributed between .70 and 1.33, and those of λ_w and λ_P in the non-linear regression between .69 and 1.00, and .75 and 1.36, respectively. The values greater than 1.00 all have rather large error estimates with them, and some of the regressions have as few as nine observations, so this large variation is not as bad as it looks. Nevertheless, it does cast some doubt upon the significance of the difference between the λ's and 1.00. Thus it seemed reasonable to test the hypothesis that λ is 1.00 for both w and P. This would mean the estimate is simply the previous period's value plus a trend term. The result of this is given as the third line of Table 3. The adjusted r-square and coefficient error estimates are not quite as good, and the Durbin-Watson statistic is noticeably worse. Nevertheless, the improvement does not seem to be enough to conclude that λ is really different from 1.00.[7] To give the reader a feeling for the degree of sensitivity to the exact period used, the next three lines of Table 3 repeat the first three, but with the period 1952–70.

Since we have concluded that the estimate of w or P is very nearly the last period's

$$P_t - E(P_{t+1} \mid t) = P_t - \lambda P_t + (1-\lambda)E(P_t \mid t-1) = (1-\lambda)[P_t - E(P_t \mid t-1)]$$

Clearly if we had used this interpretation, the coefficient of $P_t - P_t^*$, which is b_3, would have been larger by a factor of $1/(1-\lambda)$. On the other hand, the linear regression would have to be interpreted by equation (5) which has $(1-\lambda)b_3$ where equation (6) has b_3. Thus the predicted coefficient for the linear regressions would be the same. These results thus do not distinguish between equations (5) and (6) directly. Nevertheless, since $(1-\lambda)$ has different signs over different periods, the $E(P_{t+1}|t)$ interpretation would require b_3 to be changing signs the same way, clearly an unacceptable hypothesis. Thus we may regard the interpretation of P_t^* as $E(P_t|t-1)$ as being the correct one. A similar argument can be made for w_t^* if b_1 and b_2 are regarded as being equal. Certainly their ratio is small compared to $1/(1-\lambda)$, the critical ratio in the above argument.

[7] In any case, taking all of the regression results shown together it seems unlikely that λ is less than .7. Since the weight of each succeeding year in an adaptive model is decreased by a factor of $(1-\lambda)$, we may take $1/\lambda$ to be a time constant. A λ between .7 and 1.0 implies a time constant between 1.0 and 1.4 year. Clearly annual data is not well suited to make discriminations within this range.

value with a trend, there is another place adaptive expectations could be tried: the trend term. It would be reasonable to expect that term to adjust slowly to match the experienced trend. This would show up in these regressions as a change in the intercept. A test was done using a program written by Edward Prescott and Thomas Cooley. This program allows the intercept to change adaptively. The best predictions are when λ is zero for the intercept adaptation. It is not possible to quote a meaningful standard error for that estimate, but the conclusion seems to be that the existence of that type of adaptation is not supported by this evidence.

III. Discussion

It would appear that the labor supply model proposed by Lucas and Rapping (equation (1)) fits the facts of the recent *U.S.* economy fairly well. However, it should be interpreted as comparing actual wages and prices with those expected to prevail at the present time based on past information. Thus employment will differ from normal only when people are surprised by the wage or price level. This interpretation will probably require a new motivation to be developed for the equation. With that qualification, the results are quite impressive. The coefficients evaluated over the period 1930–65 agreed well with those of 1950–70. For periods including such different economic events, this is rather remarkable.

There is some trouble trying to draw definite conclusions about the expectation model. Adaptive expectations worked quite well, but λ (the weight put on the past period's actual value) had values which are nearly 1.00, and in fact probably cannot be distinguished from 1.00. No evidence was found that the expected trend moves adaptively, either, although better statistics might reverse that conclusion. The conclusion seems to be that any departures of expectations from last period's value plus a constant trend is due to long-term learning, and even then its effect has not clearly shown up here. This suggests that the Phillips curve tradeoff between inflation and unemployment may be real—if not permanent, at least over a period of years.

Before these conclusions are accepted, however, some limitations of this study should be considered. First, it uses annual data over a 21-year period. This makes the sample size too small to give strong statistical results. The consequences of using annual data for dynamic models are not clear yet, but may well be significant. Second, the 1950–70 period was rather placid economically. In trying to choose between expectations models it may well be necessary to use periods where there are fairly major events to adapt to. Somewhat related to this point is the possibility of a threshold of error in the learning of expectations: if the difference between people's expectations and reality is below this threshold, it might be assumed that no change will occur in the expectation scheme. Only when it exceeds this threshold will learning occur. If there is such a phenomenon, it will be particularly crucial to use a period with a great prevalence of surprising events. Otherwise no evidence at all may be found of learning or adaptation.

Appendix

Data Sources

All variables in the text represent logarithms of the raw data. The actual data are available from the author on request.

Lucas-Rapping Data, Extended to 1970 by Author

N = (number of persons engaged in production, Table 6.6, *Surv. Curr. Bus.*) $\times$(hours/yr by full-time employees, Denison, p. 37, normalized to 100 = 1929). The Denison series was extended beyond 1958 by regressing *BLS* weekly manufacturing hours on it (Table C-1, *Employment and Earnings*).

w = (compensation/full-time equivalent employee, Table 6.5, *Surv. Curr. Bus.*)/ (hours/year by full-time employees—as above)/P.

Q = index of effect of education on labor productivity, Dennison, p. 85. This series was extended beyond 1958 by linear extrapolation.

P = implicit price deflator on total GNP, 100 = 1958, see Table 8.1, *Surv. Curr. Bus.*

y = total GNP in 1958 dollars, Table 1.2, *Surv. Curr. Bus.*

$M=\frac{1}{L_0}\sum_i \frac{L_{0i}}{P_{0i}} P_{ti}$, where L_i is (labor force, Table A-2, *Manpower Report of the President*) and P_i is (population, *Current Population Reports*, p. 25), both in 6 age-sex groups: (14–19, 20–64, 65+)×(male, female). The subscript 0 refers to 1947–49.

BLS Data

N = total private man-hours, Table C-10, *Employment and Earnings.*

w = total private compensation/man-hour, Table C-10, *Employment and Earnings.*

Q = median years in school for labor force, Table 12, *Handbook of Labor Statistics*, interpolated graphically.

P = implicit price deflator on private GNP, 100 = 1958, Table 8.1, *Surv. Curr. Bus.*

y = private GNP in 1958 dollars, Table 1.2, *Surv. Curr. Bus.*

M = as above, but with groups (14–15, 16–17, 18–19, 20–64, 65+)×(male, female). Population is (labor force, Table A-2, *Manpower Report of the President*)+(persons not in labor force, Table A-7, *Manpower Report of the President*).

REFERENCES

T. F. Cooley and E. C. Prescott, "The Adaptive Regression Model," unpublished disc. pap., dept of economics, Univ. Pennsylvania, Nov. 1970.

E. F. Denison, *The Sources of Economic Growth in the United States and the Alternatives Before Us*, suppl. pap. no. 13, Committee for Economic Development, 1962.

M. Friedman, "The Role of Monetary Policy," *Amer. Econ. Rev.*, Mar. 1968, *58*, 1–17.

D. F. Gordon and A. Hynes, "On the Theory of Price Dynamics," in E. S. Phelps, ed., *Microeconomic Foundations of Employment and Inflation Theory*, New York 1970.

R. E. Lucas, "Econometric Testing of the Natural Rate Hypothesis," unpublished working pap., Graduate School of Ind. Adm., Carnegie-Mellon Univ. 1970.

——— and L. A. Rapping, "Real Wages, Employment, and Inflation," *J. Polit. Econ.*, Oct. 1969, *77*, 721–54.

D. T. Mortensen, "Job Search, the Duration of Unemployment, and the Philips Curve," *Amer. Econ. Rev.*, Dec. 1970, *60*, 847–62.

J. F. Muth, "Optimal Properties of Exponentially Weighted Forecasts," *J. Amer. Statist. Assn.*, June 1960, *55*, 299–306.

U.S. Bureau of the Census, *Current Population Reports*, Series P-25, Nos. 18, 21, 37, 385, 441, 448.

U.S. Bureau of Labor Statistics, *Employment and Earnings*, *17*, Washington, June 1971.

———, *Handbook of Labor Statistics 1970*, Bull. 1666, Washington 1970.

U.S. Office of Business Economics, *Surv. Curr. Bus.*, July 1971, *51*.

U.S. President, *Manpower Report of the President*, Apr. 1971.

Classical Production Functions, Technical Optimality, and Scale Adjustments of the Firm

By DALE TRUETT AND BLAINE ROBERTS*

Recently, there has been renewed interest in relations among production functions, cost curves, and scale adjustments by firms in response to factor price changes. The current literature contains a number of efforts to clear up the connection between factor prices and output adjustments of the perfectly competitive firm.[1] In particular, Charles Ferguson and Thomas Saving have related adjustments in the size of the purely competitive firm in equilibrium to the expenditure elasticity of factors, and one of the present writers has demonstrated that expenditure elasticity can be determined from factor intensity elasticity or from the slope of the expansion path in a two-factor case. A similar proof has been provided independently by Lowell Basset and Thomas Borcherding.

In part, the general lack of understanding of the precise relations between production functions, cost curves, and scale adjustments stems from neoclassical theory's divergent paths, which led to both the intuitive appeal of U-shaped cost curves and the indeterminacy of the firm in the general equilibrium model formulated by Walras and popularized by Cassel. The confusion is exacerbated by the unfamiliarity of many American economists with the excellent technical treatise of Ragnar Frisch (first published in English in 1965) and by the scant opportunities for empirical evidence of U-shaped cost curves. Alan Walters, in his well-known survey of production function and cost function literature, notes " . . . there is no large body of data which contradicts the hypothesis of a U-shaped cost curve and the fruitful results which depend upon it" (p. 52). However, empirical methods have resulted in production functions of a very restrictive nature, for example, the familiar Cobb-Douglas. In an attempt to reconcile these empirical functions with classical functions producing U-shaped cost structures, various generalizations have been made; for example, *CES*, *VES*, and recently, homothetic production functions have been estimated.

This article has several purposes. First, we develop the concept of constant homogeneity contours in the isoquant map and present a classification of nonhomogeneous production functions which clarifies the relation between production and cost curves. Second, we relate constant homogeneity contours to expenditure elasticity of factors and demonstrate that the latter is a local phenomenon and subject, in general, to reversal. Factor expenditure elasticity thus proves to have limited usefulness for determining scale adjustments when factor prices change. Finally, in the appendixes to the paper, examples of various classical production functions are presented.

I. Constant Homogeneity Contours

For simplicity, and without loss of generality, assume the production function is of the following form:

$$x = f(a, b) \tag{1}$$

where x = output, a and b are factors of production, and the function is twice differentiable in the domain of a and b. In the

* Professor of economics, University of Texas at San Antonio, and assistant professor of economics, University of Florida (on leave as research associate at Brookings Institution), respectively. This work was in part supported by a grant from the University of Florida College of Business Administration. Thanks are due Thomas R. Saving, Hajime Hori, and anonymous referees for helpful suggestions on earlier drafts.

[1] Recent work pertaining to this issue includes not only analyses of scale adjustments per se but also examinations of the determinants of input demand. Contributions include those of Lowell Bassett and Thomas E. Borcherding, D. V. T. Bear, C. E. Ferguson and T. R. Saving, Paul Mayer (1968), and Trout Rader.

relevant region, the marginal products, f_a and f_b, are positive and isoquants are smoothly convex to the origin. Furthermore, for the region relevant to our discussion, the sufficient conditions for a profit maximum are assumed:

$$f_{aa} < 0, f_{bb} < 0, \text{ and } f_{aa}f_{bb} > f_{ab}^2$$

A constant homogeneity contour (*CHC*) is the locus of all points where the production function possesses the property of homogeneity of degree k, that is,[2] where[3]

$$kx = f_b b + f_a a \tag{2}$$

The slope of a *CHC*, which is in implicit functional form in (2), can be found by differentiating (2) and holding k constant, providing such a contour exists. Thus:

$$\begin{aligned} k(f_b db + f_a da) &= f_{bb}bdb + f_{ba}bda + f_b db + f_{aa}ada \\ &\quad + f_{ab}adb + f_a da, \end{aligned} \tag{3}$$

or solving for $\left.\frac{db}{da}\right|_k$, we have:

$$\left.\frac{db}{da}\right|_k = -\frac{f_{ba}b + f_{aa}a + (1-k)f_a}{f_{bb}b + f_{ab}a + (1-k)f_b} \tag{4}$$

If the function $x=f(a, b)$ is homogeneous at every point in the isoquant map, the derivative in (4) will be undefined (i.e., $db/da=0/0$). If, however, the function produces a U-shaped long-run average cost curve, (4) will be uniquely defined for every value of a and b in the relevant region. Furthermore, as one moves along any monotonically increasing path from the origin and as successive *CHC*s are intersected, k must be first greater than one, then equal to one, and finally less than one, corresponding to falling average cost, average cost equal to marginal cost, and increasing average cost. This class of functions will be termed "classical" functions.[4]

For classical functions, (2) will be a contour. Furthermore, when $k=1$, the contour will be the locus of points that are technically optimal, in the sense that they are the points at which average returns will be maximized or average cost (for any relative factor prices) minimized. This *CHC* will be termed the *technically optimal contour* (*TOC*), and its slope is:

$$\left.\frac{db}{da}\right|_{k=1} = -\frac{f_{ba}b + f_{aa}a}{f_{bb}b + f_{ab}a} \tag{5}$$

In general, *CHC*s may have any shape, and their slope could be either positive or negative without violating the above assumptions or the sufficient conditions for a profit maximum.[5] However, there are sufficient conditions for the *CHC* to be downward sloping and to either be coincident with isoquants or to cut isoquants from above or below. For example, if $f_{ab} \leq 0$ and $k \geq 1$, then the constant homogeneity contour will be downward sloping. Following Frisch, p. 123, since the homogeneity of the function at a point corresponds to the scale elasticity at that point, the region where $k>1$ will be designated the technically preoptimal region and where $k<1$, the technically postoptimal region. In the technically preoptimal region, long-run average cost will be falling, while in the technically postoptimal region long-run average cost will be increasing.

II. Relation of the Technically Optimal Contour to Isoquants

The technically optimal contour *TOC* determines the size of the purely competitive firm in equilibrium since it traces the locus of points where average returns are equal to

[2] k is mathematically equivalent to the scale elasticity (E_s) at a particular point in the isoquant map. That is:

$$dx = f_a da + f_b db$$
$$\frac{dx}{x}x = f_a a\frac{da}{a} + f_b b\frac{db}{b}$$

if $da/a=db/b=dk/k$ and since $E_s=(dx/dk)(k/x)$ then $E_s(x)=f_a a+f_b b$ or $E_s=k$.

[3] For the n-factor case, $x=f(a_1, \ldots, a_n)$, *CHC*s are the implicit functions $kf(a_1, \ldots, a_n)=\sum_{i=1} f_i c_i$.

[4] Frisch, p. 120, describes such a function as following the *regular ultra passum* law.

[5] That is, $f_{aa}f_{bb}>f_{ab}^2$ is neither necessary nor sufficient for $db/da|_k$ to be greater than, equal to, or less than zero. Levenson and Solon, p. 107, present what is equivalent to a *TOC* in their text, *Essential Price Theory*.

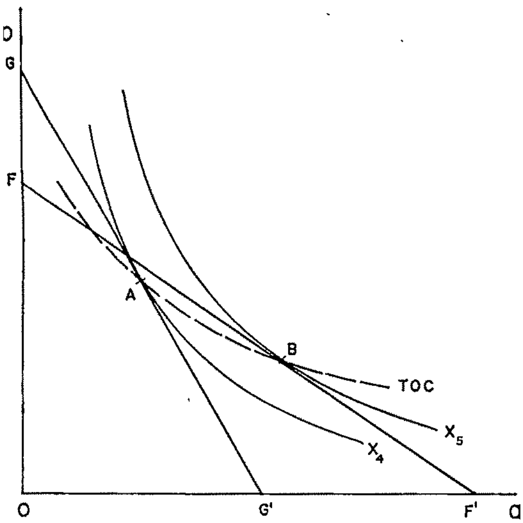


FIGURE 1. *VOS* PRODUCTION FUNCTION WHERE OPTIMAL SCALE INCREASES AS p_a/p_b FALLS

marginal returns. That is, the *TOC* traces out all points where the production function possesses the property of linear homogeneity. Thus, whether the *TOC* is coincident with or intersects isoquants from above or below will determine the change in size of the competitive firm in response to a change in factor price ratios.

It is useful to subdivide classical production functions into those with *variable optimal scale* (*VOS*) and those with *constant optimal scale* (*COS*). For *VOS* functions, the *TOC* will intersect isoquants, and in general, more than once. Figure 1 is an illustration of a variable optimal scale function where the *TOC* intersects isoquants only once. Points *A* and *B* represent the cost minimization points for factor price ratios given by the slopes of the budget lines *GG'* and *FF'*, respectively. Assuming that these are also points for which average total costs are a minimum (i.e., $k=1$), they will also lie on the *TOC*. Note that as p_a/p_b falls, optimal output increases.

It has been asserted in the literature that if a homothetic production function has an optimal scale under each price situation, then the optimal scale is the same for all price situations.[6] This is true, and in Appendix II, a proof is provided along with an example of a homothetic classical production function. Also, if *CHC*s are coincident with isoquants, then the expansion path (isocline) is a linear ray through the origin. More formally:

THEOREM 1: *If:*

$$\left.\frac{db}{da}\right|_k = \left.\frac{db}{da}\right|_{dx=0} \tag{6}$$

i.e., if

$$-\frac{f_{ab}b + f_{aa}a + (1-k)f_a}{f_{bb}b + f_{ab}a + (1-k)f_b} = -\frac{f_a}{f_b} \tag{7}$$

then

$$\sigma = \left.\frac{db}{da}\right|_{d(f_b/f_a)=0} = \frac{b}{a} \tag{8}$$

where σ is the slope of the expansion path.

PROOF:[7]

First note:

$$d\left(\frac{f_b}{f_a}\right) = \frac{f_{bb}db + f_{ab}da}{f_a} - \frac{f_b}{f_a^2}(f_{aa}da + f_{ab}db) \tag{9}$$

so that $d(f_b/f_a)=0$ implies

$$\sigma = \frac{f_b f_{aa} - f_a f_{ab}}{f_a f_{bb} - f_b f_{ab}} \tag{10}$$

[6] For example, see Meyer (1968, 1970). Meyer (1968) provides an example of a homothetic function, but it does not produce a U-shaped cost curve.

[7] This theorem can be proved for the n-factor case by beginning as follows. A point of tangency of the *TOC* and an isoquant is a solution to the system of homogeneous equations $\sum_i \sum_j f_{ij} a_j dx_i = 0$ and $\sum_i f_i dx_i = 0$. Thus if

$$\frac{\sum_j f_{ij}(\langle\bar{a}\rangle)a_j}{f_i(\langle\bar{a}\rangle)} = \frac{\sum_j f_{kj}(\langle\bar{a}\rangle)a_j}{f_k(\langle\bar{a}\rangle)} \quad \text{for all } i, k = 1, \ldots, n$$

holds for every vector of inputs $\langle\bar{a}\rangle$ such that $f(\langle\bar{a}\rangle) = \bar{y}$ (a constant), then the *TOC* is coincident with an isoquant.

From (7),

$$(11)\quad f_b f_{ab} b + f_b f_{aa} a + (1-k) f_b f_a = f_a f_{bb} b + f_a f_{ab} a + (1-k) f_b f_a$$

Thus,

$$(12)\quad \frac{f_b f_{aa} - f_a f_{ab}}{f_a f_{bb} - f_b f_{ab}} = \frac{b}{a}$$

It can easily be shown that the left-hand term of (12) is the slope of an isocline.[8] Also, any function with *CHC*s coincident with isoquants is homothetic. That is, if $x=f(a, b)$ and if the premises of Theorem 1 are true, then there exists some monotonic transformation, g, of a linearly homogeneous function $h(a, b)$ such that $x=g[h(a, b)]$. However, it does not necessarily follow that all *COS* functions are homothetic. That is, a production function may have its *TOC* coincident with an isoquant while other *CHC*s ($k \neq 1$) intersect isoquants.

III. The *VOS* Function and Factor Expenditure Elasticity Reversal

For a *VOS* production function, the *TOC* must intersect isoquants, and within the range of positive factor prices it *may* intersect some isoquants more than once. Figure 2 illustrates a case in which the *TOC* intersects an isoquant twice. If relative factor prices are given by the slope of GG', optimal long-run output will be x_3. Any fall in relative factor prices will result in an optimal output greater than x_3 as long as factor prices do not fall below the ratio given by FF'. If the factor price ratio does fall below OF/OF', optimal long-run output will be less than x_3. Thus from point C a fall in p_a, for example, could cause optimal output *either* to increase

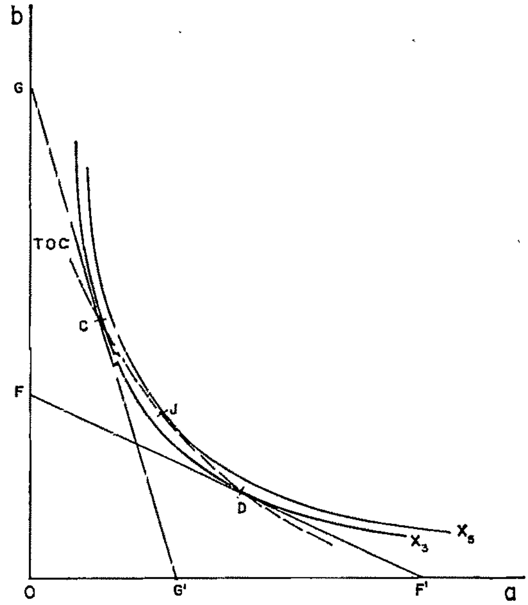


FIGURE 2. *VOS* PRODUCTION FUNCTION WITH A MAXIMUM OPTIMAL OUTPUT AND EXPENDITURE ELASTICITY REVERSAL

or to decrease depending upon the magnitude of the price change. Further, in Figure 2 at J there exists a factor price ratio which will yield the *maximum* optimal output in the long run. Thus, if a *VOS* function has a point such as J within the range of possible factor prices, the perfectly competitive firm will not produce a long-run output greater than the level consistent with that point. In Appendix I we provide an example of a *COS* production function with a maximum optimal scale.[9]

From the foregoing, it is evident that there is no general relation between a change in the price of a given factor and adjustment of optimal output for a *VOS* production function. Optimal output may either rise or fall when a factor price changes, depending upon the exact *form* of the production function and the *magnitude* of the change in price. Indeed, when a long-run equilibrium occurs at a point of tangency between the *TOC* and an

[8] For $x=f(a, b)$ and from the necessary conditions for minimum cost where $c=ap_a+bp_b$ is total cost, let

$$Z = \frac{f_a}{f_b} - \frac{p_a}{p_b} = 0, \quad \text{then}$$

$$\frac{dZ}{da} = \frac{f_{aa}}{f_b} + \frac{f_{ab}}{f_b}\frac{db}{da} - \frac{f_a f_{ab}}{f_b^2} - \frac{f_a f_{bb}}{f_b^2}\frac{db}{da} = 0$$

Solving for db/da, we obtain

$$\frac{db}{da} = f_b f_{aa} - f_a f_{ab} / f_a f_{bb} - f_b f_{ab}$$

[9] The reverse case of Figure 2 is also possible. The *TOC* may intersect isoquants twice and be tangent to one which is a *minimum* output for the perfectly competitive firm in long-run equilibrium.

isoquant such as J in Figure 2, *any* change in relative factor prices will reduce long-run optimal output.

In their article on scale adjustments, Ferguson and Saving conclude that when a single input price changes, adjustments in long-run optimum output depend upon the expenditure elasticity of the price-changing input. Let

$$S_a = \frac{c}{a}\,\frac{da}{dc} \tag{13}$$

be the expenditure elasticity of input a, where

$$c = ap_a + bp_b \tag{14}$$

is total cost. Assuming $S_a > 0$, factor a is normal, neutral, or superior as $S_a \lesseqgtr 1$. In the event that $S_a < 0$, factor a is inferior.

Ferguson and Saving state:

> [The] long-run equilibrium output of a perfectly competitive firm must always vary directly with the price of an inferior or normal input and must always vary inversely with the price of a superior input. [p. 777]

The above is correct, but it is necessary to indicate that the expenditure elasticity of a factor may *reverse* over the range of an observed price change, thereby rendering impossible the prediction of scale adjustment from expenditure elasticity and direction of factor price change.

For normal factors of production, we can, however, determine the relation of expenditure elasticity to isoquants and the technically optimal contour. Specifically, at any technically optimal point for a downward sloping TOC, if the slope of the TOC is greater than the slope of an intersecting isoquant, then the expenditure elasticity of factor a is greater than one. Recalling from (8) that σ denotes the slope of the expansion path, it can be shown that if:

$$\left.\frac{db}{da}\right|_{k=1} > -\frac{f_a}{f_b} \tag{15}$$

then[10]

$$\sigma < \frac{b}{a} \tag{16}$$

Let the factor intensity elasticity (elasticity of expansion) be

$$E_L = \frac{\sigma}{b/a} \tag{17}$$

as defined previously in a paper by Truett. From the foregoing, the numerator of (17) is the slope of the expansion path, while the denominator is the factor ratio. If $E_L < 1$, production becomes more a-intensive as the firm expands at given factor prices; the converse is true if $E_L > 1$. In the former case the proportion of total outlay spent for a increases, while that spent for b falls. Thus, $S_a > 1 > S_b$. This can be easily shown. Since:

$$S_a = \frac{c}{a}\,\frac{da}{dc} = \frac{p_a + p_b(b/a)}{p_a + p_b\sigma} \tag{18}$$

S_a must be greater than one if $E_L < 1$. Thus, at any point where a downsloping TOC cuts an isoquant from below, factor a is superior while factor b is normal. This is the case at points A and B in Figure 1 and at point C in Figure 2. However, by similar reasoning when the TOC cuts an isoquant from *above* (such as point D in Figure 2), $E_L > 1$ and

[10] Since $-\dfrac{f_{ab}b + f_{aa}a}{f_{bb}b + f_{ab}a} < 0$,

either: or:

I. $f_{bb}b + f_{ab}a < 0$, $f_{ab}b + f_{aa}a < 0$

II. $f_{ab}b + f_{aa}a > 0$, $f_{bb}b + f_{ab}a > 0$

II. implies $f_{aa}f_{bb} < f_{ab}^2$ and violates the sufficient conditions for a profit maximum. Therefore, using I. and if

$\left.\dfrac{db}{da}\right|_{k=1} > -\dfrac{f_a}{f_b}$, then $(f_{aa}f_b - f_{ab}f_a)a > b(f_af_{ab} - f_bf_{ab})$.

Since

$$\sigma = \frac{(f_bf_{aa} - f_af_{ab})}{(f_af_{bb} - f_bf_{ab})} > 0,$$

the numerator and denominator must be of the same sign. Therefore, both numerator and denominator must be negative for the sufficient conditions for a profit maximum to hold and, in particular, $f_af_{bb} - f_bf_{ab} < 0$. Consequently, it follows that $\sigma < b/a$.

$S_b > 1 > S_a$. In fact, it is evident that J in Figure 2 is a point of expenditure elasticity reversal, since a fall in p_a will cause optimal output to increase from higher points along the *TOC*, but will cause optimal output to fall below J. At this juncture, we can state:

THEOREM 2: *If:*

$$(19) \qquad \left.\frac{db}{da}\right|_{k=1} < 0; \qquad \sigma > 0; \quad and$$

$$\left.\frac{db}{da}\right|_{k=1} \gtrless -\frac{f_a}{f_b};$$

then:

$$(20) \quad E_L \lesseqgtr 1 \quad and \quad S_a \gtreqless 1 \quad while \quad S_b \lesseqgtr 1;$$

which is proved by the foregoing argument.[11]

Also, at points such as A and B in Figure 1 (C in Figure 2), the expansion path (not shown in figures) is less steep than a factor-ray through the origin, since $\sigma < b/a$. In other words, where the *TOC* intersects isoquants from *below*, the slope of isoclines will be less than the slope of a ray through the origin; the reverse will be true if the *TOC* intersects isoquants from *above* (at D in Figure 2, for example). Since ridge lines are isoclines, this suggests that the "textbook" nonhomogeneous function with inward-bending ridge lines (see James Henderson and Richard Quandt, p. 48, or Albert Levenson and Babette Solon, p. 107) is characterized by expenditure elasticity reversal of factors and isoquants which intersect the *TOC* more than once. Appendix I provides an example of such a function.

IV. Summary and Conclusions

Classical production functions are those which produce U-shaped long-run average cost curves. When such functions are examined in terms of contours for constant homogeneity, as output increases the degree of homogeneity, k, is initially greater than one but decreases and becomes less than one. Where $k=1$, marginal and average returns are equal, and points of technical optimality occur on the function. Regardless of factor prices, minimum long-run average cost will correspond to some factor combination on the locus of technically optimal points, the technically optimal contour. If the *TOC* is a single isoquant, the production function will have a constant optimal scale, and optimal output will be the same for all factor prices. On the other hand, when the *TOC intersects* isoquants, a change in relative factor prices will cause optimal output to increase or decrease. If the *TOC* cuts each isoquant only once, the change in optimal output following a factor price change will depend upon the direction of the price change and whether the *TOC* cuts isoquants from above or below. In this case, the direction of scale adjustment can be unambiguously predicted from the expenditure elasticity of the price-changing factor. In the event that the *TOC* intersects isoquants twice or more, *magnitude* as well as direction of factor price change is important to the outcome. This final case is characterized by expenditure elasticity reversal. Here, factor expenditure elasticities are meaningful only in a local context and provide no clue to the perfectly competitive firm's scale adjustment when a finite factor price change occurs.

APPENDIX

I. Nonhomothetic Classical *VOS* Function with Expenditure Elasticity Reversal

A variable optimal scale function similar to that of Figure 2 above can be specified by adding internal economies and diseconomies to a linearly homogeneous production function. The function introduced by Ryuzo Sato in a series of exchanges some years ago in this *Review*, provides a precise textbook example of the linearly homogeneous production function, since it produces a "Knightian" total product curve for the short-run variable factor. Although it is desirable heuristically that the classical *VOS* function have short-run properties similar to the Sato function, the *VOS* function, by definition, will be nonhomogeneous.

The following modified Sato function will fit the *VOS* function in Figure 3:

[11] This theorem also can be generalized to the n-factor case in a straightforward manner.

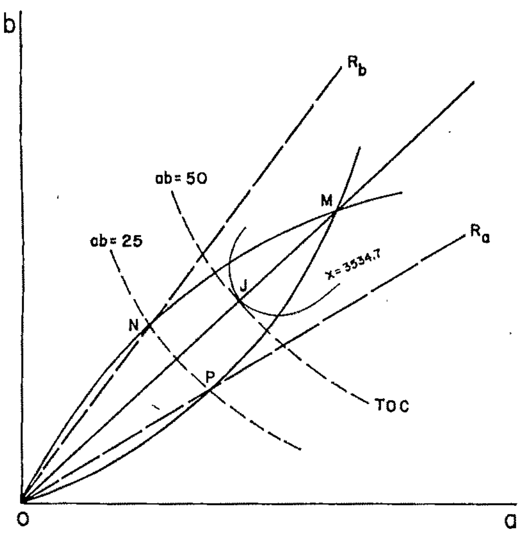


FIGURE 3. MODIFIED SATO FUNCTION WITH VARIABLE OPTIMAL SCALE AND EXPENDITURE ELASTICITY REVERSAL

$$\text{(I.1)} \qquad \frac{(ab)^2}{Da^3 + Eb^3} + (Fab - Ga^2b^2)$$

where $D=E=.001$, $F=.5$, and $G=.01$. The disturbance term $(Fab-Ga^2b^2)$ can be viewed as the combined effect of internal economies and diseconomies, an effect which will be positive at low values of ab but negative when ab is relatively large. The combined disturbances will equal zero when

$$\text{(I.2)} \qquad ab = F/G = 50$$

or along a rectangular hyperbola in the isoquant map. At any point on this contour (the TOC), the function will behave as the original Sato function. Further, when $F/2G = ab = 25$, there will be no disturbance to the Sato marginal products; thus the original Sato ridge lines (R_a and R_b) cross those for the modified function where the rectangular hyperbola $ab=25$ also intersects the latter.

At the point of maximum optimal scale, the absolute slope of an isoquant must equal that of a factor ray. For the parameters specified above, this occurs when:

$$\text{(I.3)} \qquad b/a = \sqrt[3]{D/E} = 1$$

However, for all values of $b/a=1$ between the origin and M, Theorem 1 holds. That is, if a CHC is tangent to an isoquant for the modified Sato function, then:

$$\text{(I.4)} \quad b/a = \sqrt[3]{D/E} = 1 \qquad \text{if } D = E$$

It follows that tangencies between CHCs and isoquants occur all along OM in Figure 3. In the relevant range above OM, all isoclines are shaped like ridge line ONM; thus $S_a > 1 > S_b$ and CHCs cut isoquants from below. Below OM, isoclines are shaped like ridge line OPM, and $S_b > 1 > S_a$ while CHCs cut isoquants from above. It is evident that all points of expenditure elasticity reversal occur along OM, and for the parameters specified above, maximum optimal scale occurs at J where $a=b=7.07$. and output is 3,534.7 units of x.

II. Homothetic Classical Production Functions

Intuitively, a homothetic classical production function must have constant optimal scale because the transformation f in $x=f[h(a, b)]$ where h is homogeneous only "relabels" isoquants but does not change their shape. In other words, if the transformation f creates technical optimality for $\bar{a}, \bar{b}$, then it must do so for all $h(a, b)=h(\bar{a}, \bar{b})$. A formal proof and example follow.

THEOREM 3: *A homothetic production function has constant optimal scale; that is, the technically optimal contour is coincident with an isoquant.*

PROOF:

By definition, x is homothetic if $x=f[h(a, b)]$ where $h(a, b)=ah_a+bh_b$. Therefore, the technically optimal contour is

$$\left.\frac{db}{da}\right|_{k=1} = -\frac{f'bh_{ba} + bh_bh_af'' + f'ah_{aa} + af''h_a^2}{f'bh_{bb} + bf''h_b^2 + f'ah_{ba} + ah_ah_bf''}$$

Since h is homogeneous,

$$ah_{aa} + bh_{ab} = bh_{bb} + ah_{ab} = 0$$

$$\left.\frac{db}{da}\right|_{k=1} = -\frac{h_b h_a b + h_a^2 a}{h_b^2 b + h_a h_b a} = -\frac{h_a}{h_b}$$

An example of a homothetic classical production function is

$$x = a^\alpha b^\beta - ca^{2\alpha} b^{2\beta}$$

where $\alpha > 0$, $\beta > 0$, $\alpha + \beta > 1$, $c > 0$, and the function is defined for the region where $a \geq 0$, $b \geq 0$, and $a^\alpha b^\beta \leq 1/c$. If we let $a^\alpha b^\beta = f$, then $x = f - cf^2$, and f is a Cobb-Douglas function with homogeneity of $\alpha + \beta$. Since

$$\frac{x_a}{x_b} = \frac{\alpha a^{\alpha-1} b^\beta - 2\alpha c a^{2\alpha-1} b^{2\beta}}{\beta a^\alpha b^{\beta-1} - 2\beta c a^{2\alpha} b^{2\beta-1}} = \frac{\alpha b}{\beta a}$$

the function is clearly homothetic. The slope of the *TOC* is:

$$\left.\frac{db}{da}\right|_{k=1}$$

$$= -\frac{\alpha b}{\beta a} \frac{[\alpha + \beta - 1 - 2ca^\alpha b^\beta (2\beta + 2\alpha - 1)]}{[\alpha + \beta - 1 - 2ca^\alpha b^\beta (2\alpha + 2\beta - 1)]}$$

$$= -\frac{\alpha b}{\beta a}$$

Thus, the *TOC* is indeed coincident with an isoquant.

REFERENCES

L. Bassett and T. E. Borcherding, "The Relationship Between Firm Size and Factor Price," *Quart. J. Econ.*, Aug. 1970, *84*, 518–22.

D. V. T. Bear, "Inferior Inputs and the Firm," *J. Polit. Econ.*, June 1965, *73*, 287–89.

S. Clemhout, "The Class of Homothetic Isoquant Production Functions," *Rev. Econ. Stud.*, Jan. 1968, *35*, 139–65.

C. E. Ferguson and T. R. Saving, "Long-Run Scale Adjustments of a Perfectly Competitive Firm and Industry," *Amer. Econ. Rev.*, Dec. 1969, *59*, 774–83.

R. Frisch, *Theory of Production*, Dordrecht 1965.

J. M. Henderson and R. E. Quandt, *Microeconomic Theory*, New York 1958.

A. M. Levenson and B. Solon, *Essential Price Theory*, New York 1971.

P. A. Meyer, "An Aggregate Homothetic Production Function," *Southern Econ. J.*, Jan. 1970, *36*, 229–38.

———, "A Paradox on Profits and Factor Prices: Reply," *Amer. Econ. Rev.*, Sept. 1968, *58*, 923–43.

T. Rader, "Normally Factor Inputs are Never Gross Substitutes," *J. Polit. Econ.*, Jan. 1968, *76*, 38–43.

R. Sato, "Diminishing Returns and Linear Homogeneity: Comment," *Amer. Econ. Rev.*, Sept. 1964, *54*, 744–45.

D. Truett, "The Elasticity of Expansion and Scale Adjustments of the Firm," *Amer. Econ. Rev.*, Dec. 1971, *61*, 962–65.

A. A. Walters, "Production and Cost Functions: An Econometric Survey," *Econometrica*, Jan.-Apr. 1963, *31*, 1–66.

Peasants, Procreation, and Pensions: Note

By Clive Southey*

In his recent article in this *Review*, Philip Neher attempts to demonstrate rigorously that there is substance to the allegation that underdeveloped nations frequently have unduly high birth rates in order to satisfy the "pension motive," that is to say, parents have children to guarantee that someone will provide for them in their old age.

Neher's analysis is fairly general but he feels that the case which "particularly merits attention" is that in which egoistic parents care only for their own utility, having little regard for the world subsequent to their own deaths. This case exemplifies " . . . the *fundamental market failure* which exists when the consequence of current decisions carry beyond the decision makers' planning horizon" (p. 386). At this point, Neher adds in a footnote: "Note that this occurs independent of 'like father like son' effects. This model takes the fertility of future generations as exogenous" (fn. 16, p. 386).

In this footnote Neher uncovers a basic deficiency of his model, for to suggest that the number of grandchildren I have will be independent of the number of children I produce, is too unrealistic and particularly so in an economy which achieves a steady state. To note as Neher does that "There seems no very good way to make future fertility endogenous . . . " (fn. 7, p. 382) is not sufficient justification for making these decisions exogenous. This paper demonstrates that if a steady state were to prevail, such that succeeding generations learn that a "like father like son" mentality was prevailing, *there might be no market failure precisely in the case of egoistic parents.*

To demonstrate this we follow Neher's model and notation closely and confine ourselves to the case where $U=U(C_1)+U(C_2)$, so that although equal weight is given to adulthood and grand-parenthood $(D=1)$ future generations and even the utility of one's children while alive matter not at all $(E=0)$. If parents believe that their children will make the same decisions as they do, a belief that is not necessarily true, then Neher's equation (2) remains unchanged as follows:

$$C_1 = \frac{\overline{X}}{P_1} = \frac{f(\overline{F}_1)}{S_1 + \overline{F}_1 + \overline{G}_1} \tag{2}$$

However, his equation (3) under his simplifying assumption of certainty of survival will now read:

$$C_2 = \frac{X_2}{P_2} = \frac{f(S_1)}{2S_1 + \overline{F}_1} \tag{3}$$

It follows that Neher's equations (5) and (6) will become:

$$\frac{dC_1}{dS_1} = -\frac{1}{P_1}C_1 = -\frac{1}{P_1}\cdot\frac{f(\overline{F}_1)}{P_1} \tag{5}$$

and

$$\frac{dC_2}{dS_1} = \frac{(2S_1 + \overline{F}_1)f'(S_1) - 2f(S_1)}{(2S_1 + \overline{F}_1)^2} \tag{6}$$

Now according to Neher, in an economy obeying the "Golden Rule of Fertility" so that the population enjoys maximum sustainable level of per capita income, the average product of labor $f(S_1)/S_1$ is equal to the wage rate which is also equal to the marginal product of labor: $w=f'(S)$. Furthermore with a steady-state population $G_1=F_1=S_1=S_2$.

Thus substituting in (5) and (6) we obtain

$$\begin{aligned}\frac{dC_2}{dC_1} &= \frac{-(3S_1)^2}{f(S_1)}\cdot\frac{3S_1\cdot w - 2f(S_1)}{(3S_1)^2} \\ &= \frac{-3f(S_1) + 2f(S_1)}{f(S_1)} \\ &= -1\end{aligned}$$

* Assistant professor, University of Guelph.

This, however, is precisely the condition cited by Neher for nonviolation of the Golden Rule (p. 385). Hence, we have proven that the pension motive is consistent with the Golden Rule in this instance.

The above analysis is simply a statement about the possible existence of an equilibrium at the Golden Rule. It remains to consider whether this equilibrium is stable. We do not present an exhaustive treatment of this difficult problem, but the following analysis suggests that if the system is disturbed by say an exogenous technical change, it will move towards the new Golden Rule population.

We continue to assume that parents believe their children will simply reproduce themselves. Although this belief will be clearly contradicted by experience during any adjustment process, it would still appear to be more plausible than Neher's assumption that the number of grandchildren is determined exogenously.

If the economy is initially in a steady state with certainty of survival then $\overline{F}_1=\overline{G}_1$ so that with egoistic parents, the first-order conditions for the optimum as in (5) and (6) above can be reduced to

$$(7) \quad \frac{f(\overline{F}_1)}{(S_1+2\overline{F}_1)}=\frac{(2S_1+\overline{F}_1)f'(S_1)-2f(S_1)}{(2S_1+\overline{F}_1)}$$

If $S_1 \gtreqless F_1$ then $(2S_1+\overline{F}_1) \gtreqless (S_1+2\overline{F}_1)$ and thus by examining the differences in magnitude of the two numerators in (7) we can ascertain whether changes in population are consistent with the stability hypothesis. If (7) is satisfied, it follows that for consistency,

$$(8) \qquad S_1 \gtreqless F_1$$

$$\text{if} \quad (2S_1+\overline{F}_1)f'(S_1)-2f(S_1)-f(\overline{F}_1) \gtreqless 0$$

Consider technical changes such that the new Golden Rule population exceeds that prevailing at present. It follows that the average productivity of labor at existing population levels and for small increments therein, is less than the marginal product $f(S_1)/S_1<f'(S_1)$.

Rearranging terms in (8) and dividing through by S_1 we get

$$(9) \qquad S_1 \gtreqless F_1 \qquad \text{if} \left\{2f'(S_1)-\frac{2f(S_1)}{S_1}\right\} + \left\{\frac{\overline{F}_1}{S_1}f'(S_1)-\frac{f(\overline{F}_1)}{S_1}\right\} \gtreqless 0$$

The proof that in fact population will expand follows by contradiction. Suppose that decisions are made such that $S_1<F_1$; the first term of (9) is always positive. The sign of the second term is invariant when multiplied by $S_1/\overline{F}_1$ which yields

$$f'(S_1)-\frac{f(F_1)}{F_1}$$

and a sufficient (but not necessary) condition for this to be positive is that $f''(S_1)$ is negative. Since this is certainly true for small deviations from the Golden Rule population, it follows that the entire expression in (9) is positive; this however, contradicts the first-order conditions for population constancy or decrease and thus we must conclude that in fact population will expand. A similar argument follows for the case when the technical change calls for a decrease in Golden Rule population, and thus stability follows: not only is a share-alike system with egoistic parents consistent with the Golden Rule but it will also achieve that rule as long as parents subscribe to the belief that their children will behave in a "like father like son" manner.

By making future populations independent of the number of children born, Neher has only allowed for a "one-way" externality. What he therefore fails to incorporate is the fact that without malice, the children of peasants will in turn impose an externality on their parents by also having children and so reducing the grandparents' share; the absence of private pension schemes and the share-alike mentality is a double-edged sword that cuts both ways. It is of particular interest that in this one instance of egoistic parents, the result is compatible with the so-called Golden Rule of Fertility. Clearly this result is dependent on the specific assumptions made ($p=g=D=1$, and $E=0$ in Neher's model) as well as the as-

sumption that grandchildren claim as much of the output as do grandparents. That these *ad hoc* assumptions happen to be compatible with the Golden Rule should not detract from the potential that Neher's model has for further investigation of the role of institutions on fertility decisions. Nevertheless, at this juncture there is no a priori theoretical justification of the pension motive being a cause of "over population."

REFERENCES

P. Neher, "Peasant, Procreation and Pensions," *Amer. Econ. Rev.*, June 1971, *79*, 380–89.

Stochastic Dominance, Efficiency Criteria, and Efficient Portfolios: The Multi-Period Case

By HAIM LEVY*

The theoretical and the empirical work in the theory of investment decision making under conditions of uncertainty (particularly in portfolio-selection theory) has been done so far under the very restricting assumption that all investors have a single target date for wealth accumulation, or investment horizon. They buy their portfolios on the same date with the object of maximizing their expected utility from terminal wealth.[1] The papers of James Tobin and Jan Mossin who deal with the multi-period mean-variance criterion are the exception.

The assumption of a single target date for wealth accumulation is unrealistic since investors usually have different planning horizons. Some tend to favor short-term investments, while others (particularly institutional investors) tend to invest in long range ventures.[2]

The purpose of this paper is to analyze the relationship between the efficient set of portfolios constructed for different investment horizons. More specifically, we assume that all investors of a given class (for example, risk averters) can choose between only two portfolios with well-defined distributions F_1 and G_1, whose one-period return is given by the random variables X_1^F and X_1^G, respectively. Now, if all investors have the same investment horizon, for example, one year, then they are confronted with the same set of distributions, and hence they will all choose the same efficient set. However, if the investment horizon of some investors is, say, two years, these investors are confronted with distributions $F \equiv F_1F_2$ and $G \equiv G_1G_2$, derived from the one-period distributions.[3] The return on the two-year portfolios are given by $X_1^FX_2^F$ and $X_1^GX_2^G$, respectively, and obviously, the distribution of X_1^F (or X_2^F) is not identical to the distribution of $X_1^FX_2^F$.[4] Consequently, although all investors are confronted with the same set of one-period available portfolios, this set is completely different from the available set of portfolios of the multi-period investors. Since the efficient set of portfolios is a subset of all available portfolios, one would expect the efficient set to be a function of the investment horizon—i.e., investors for one year might decide that portfolio F_1 is included in the efficient set and G_1 is not, while two-period (or multi-period) investors might reach the opposite decision.

In this article we analyze the role of the investment horizon in determining the efficient set. Apart from the obvious theoretical importance of the above analysis, it may also have some practical merit: suppose that

* University of California, Berkeley and the Hebrew University, Jerusalem. I wish to acknowledge the helpful comments of Fred Arditti, Nils Hakansson, Jacob Paroush, Menachem Yaari, Meir Schneller, and a referee. Part of the research was supported by a grant from Dean Witter Foundation, and from the Faculty of Social Science, Hebrew University.

[1] See, for example, the books of Harry Markowitz, Haim Levy and Marshall Sarnat, and articles by William Sharpe, John Lintner, Michael Rothschild and Joseph Stiglitz.

[2] In the conventional model of the theory of investment there is no constraint on the length of the period.

[3] We assume that the returns in period 1 and in period 2 are independent. Notice, also, that we assume that the investor chooses the same investment for the whole period; hence, in this model, portfolios F_1G_2, G_1F_2 are not permissible. This means that at the time that the decision is made, the investor does not intend to switch portfolios. Of course, after the first period is over the investor realizes the first-period return and he may change his estimates as well as his portfolio. In this case, the whole process of finding the efficient set starts again at the beginning of the second period.

[4] Moreover, in the general case, the distribution of X_1^F and the distribution of $X_1^FX_2^F$ do not even belong to the same statistical family. For example, if X_1^F and X_2^F have identical normal distributions, the random variable $X_1^FX_2^F$ is not distributed according to the Normal Law.

an investment consultant tries to screen all available portfolios—i.e., to build the efficient portfolio set for his customers. What will be the impact of various investment horizons on his final advice? Should he construct the same efficient set for all his customers, or should he build different efficient sets for investors with different investment horizons?[5]

Section I of this study gives a short review of two efficiency criteria that have been developed for the one-period case. In Sections II and III we analyze the two multiperiod efficiency criteria. Concluding remarks are given in Section IV.

I. Efficiency Criteria—The One-Period Case

An efficiency criterion is a rule with which we divide all the available portfolios into two groups—the efficient set and the inefficient set; for every portfolio included in the inefficient set, there is at least one preferable portfolio in the efficient set. Obviously, we have more than one criterion, and by changing the assumptions concerning the investors' preferences, one might get several efficiency rules (and, hence, several efficiency sets). However, given a set of assumptions on the investors' utility function, the criterion that minimizes the size of the efficient set is the *optimal* efficiency criterion.

A. *General Efficiency Criterion*

In the most general case, where investors are assumed to have no systematic preferences with respect to risk (i.e., no restrictions are placed on investors' utility function beyond the assumption that their first derivatives be nonnegative), it has been proved (see articles by James Quirk and Rubin Saposnik, Joseph Hadar and William Russell, and Giora Hanoch and Levy) that option F eliminates option G from the efficient set (i.e., F dominates G, or FDG) for all (utility) functions $\phi(x)$ (where $\phi'(x)$ is positive and continuous throughout the domain of x) if, and only if, $F(x) \leq G(x)$ for all values of x, and $F(x_0) < G(x_0)$ for some value x_0, where F and G denote the cumulative distributions of the two options under consideration. That is to say,

(1) $F(x) \leq G(x)$ for all values of x

and $F(x_0) < G(x_0) \Leftrightarrow$

$\Delta E\phi \equiv E_F\phi(x) - E_G\phi(x) \geq 0$, for all $\phi(x)$ and for some nondecreasing utility function $\Delta E\phi > 0$.

B. *Efficiency Criterion in the Case of Risk Aversion*

Under conditions of risk aversion—i.e., when one assumes a nondecreasing concave utility function ($\phi'(x) > 0$, $\phi''(x) < 0$ and $\phi'(x)$ and $\phi''(x)$ are continuous throughout), it has been proved that option F dominates option G (i.e., F eliminates G from the risk averters efficient set) if, and only if, $\int_{-\infty}^{x} [G(t) - F(t)]dt \geq 0$ for all values of x (and $G \neq F$ for some value x_0). This rule can be written as:

(2) $$\int_{-\infty}^{x} [G(t) - F(t)]dt \geq 0 \text{ for every } x$$

with a strong inequality for some $x_0 \Leftrightarrow$

$\Delta E\phi \equiv E_F\phi(x) - E_G\phi(x) \geq 0$, for every $\phi(x)$, with $\phi'(x) > 0$, $\phi''(x) < 0$, and for some nondecreasing concave utility function $\Delta E\phi > 0$.

For the purpose of this paper, another version of (2) should be developed. The right side of (2) is, by definition,

$$\Delta E\phi \equiv E_F\phi(x) - E_G\phi(x) = \int_{-\infty}^{\infty} [f(t) - g(t)]\phi(t)dt$$

where f and g stand for the density functions of the two options, respectively. Integrating by parts yields,

$$\Delta E\phi \equiv \int_{-\infty}^{\infty} [f(t) - g(t)]\phi(t)dt = \{[F(x) - G(x)]\phi(x)\}_{-\infty}^{+\infty} + \int_{-\infty}^{\infty} [G(t) - F(t)]\phi'(t)dt$$

[5] We shall assume, as in the one-period case, that the investor does have a certain future point of time in which he is willing to maximize the utility of his wealth. But unlike the one-period case, the horizons should not be identical for all investors.

Since the first term on the right side vanishes[6] we can replace (2) by (2′)

$$\int_{-\infty}^{x}[G(t) - F(t)]dt \geq 0 \tag{2′}$$

for every x (with strong inequality for some x)$\Leftrightarrow$

$$\int_{-\infty}^{\infty}[G(t) - F(t)]\phi'(t)dt \geq 0$$

for every nondecreasing concave $\phi(x)$.

We shall use these results in the following sections.

II. Multi-Period Dominance in the Case of an Unrestricted Utility Function

In the rest of the paper we shall prove several theorems, first for the two-period horizon, and then the results will be generalized for an n-period horizon. Like Tobin, we assume independence over time. More precisely, we accept the Random Walk Theory (see, for example, Paul Cootner).

Let R_i denote the return on one dollar invested in period i. The final wealth for an n-period horizon is therefore given by $1+R=\prod_{i=1}^{n}(1+R_i)$. Denoting[7] $1+R_i\equiv x_i$, the joint density function of $\prod_{i=1}^{n}x_i$ is $f(x_1, \ldots, x_n)$. Accepting the Random Walk hypothesis, the joint density function can be written as $f(x_1, \ldots, x_{n-1})f(x_n)$, since $f(x_n)$ is independent of $f(x_1, \ldots, x_{n-1})$. For the two-period case, independence implies $f(x_1, x_2)=f(x_1)f(x_2)$.

THEOREM 1: *A sufficient condition for a two-period risk $F(x)$ to dominate another two-period risk $G(x)$, for all nondecreasing utility functions, is:*[8] *F_iDG_i $(i=1, 2)$, where F_i and G_i denote the single period distributions.*

PROOF:

Let x_1 and x_2 be two continuous random variables for periods 1 and 2, respectively, and $x=x_1x_2$ is the random variable (return) for the two periods. Since F_iDG_i $(i=1, 2)$, it is given that:

$$F_1(x_1) \leq G_1(x_1) \quad \text{for every } x_1 \tag{3}$$
$$\text{and } F_1(x_{1_0})<G_1(x_{1_0}) \quad \text{for some } x_{1_0}$$

$$F_2(x_2) \leq G_2(x_2) \quad \text{for every } x_2 \tag{4}$$
$$\text{and } F_2(x_{2_0})<G_2(x_{2_0}) \quad \text{for some value } x_{2_0}$$

Assuming independence over time, the distribution function of the *two-period* return is given by:

$$\begin{aligned} F(x) &= P_r(x \leq x) \\ &= \int_0^{\infty}\int_0^{x/x_1} f_1(t_1)f_2(t_2)dt_2dt_1 \\ &= \int_0^{\infty} F_2(x/t_1)f_1(t_1)dt_1 \end{aligned} \tag{5}$$

Similarly, the distribution function of option G is given by:

$$G(x) = \int_0^{\infty} G_2(x/t_1)g_1(t_1)dt_1, \tag{6}$$

where f_i, g_i stand for the density functions of the two risks in period i.

We must prove that the two-period distribution F, dominates the two-period distribution G, i.e., $F(x)\leq G(x)$ for every x, and $F(x_0)<G(x_0)$ for some value x_0. Using (5) and (6) one has to prove that (7) holds,

$$\int_0^{\infty} F_2(x/t_1)f_1(t_1)dt_1 \leq \int_0^{\infty} G_2(x/t_1)g_1(t_1)dt_1 \tag{7}$$

[6] If $F(x)$ and $G(x)$ are finite, this term obviously equals zero. If $F(x)$ and $G(x)$ are infinite (for example, have normal distributions), the proof that this term equals zero is a bit more complicated.

[7] It is clear from the above definition that $0\leq x<\infty$, the lowest possible return in each period is $R_i=-1$, i.e., the stock price drops to zero. Obviously, if we have a distribution with $P_r(x_i=0)=1$, this portfolio is excluded from the efficient set and we shall not deal below with such degenerate distributions. By the same argument, we can assume that $P_r(x\geq M)=0$, where M is a large number.

[8] F dominates G (or FDG) means that every investor in a given class (e.g., all nondecreased utility functions) prefers option F over option G, and hence G can be eliminated from the efficient set. In other words FDG implies that $E_F\phi(x)\geq E_G\phi(x)$ for all $\phi(x)$ in the class under consideration and for some $\phi(x)$ a strong inequality holds.

with a strict inequality for some value x_0.

Condition (4) can be rewritten as:

$$G_2(x/t_1) = F_2(x/t_1) + \delta(x/t_1)$$

where $\delta(x/t_1) \geq 0$ and for some value x_0/t_{1_0}, $\delta(x_0/t_{1_0}) > 0$

Substituting this result in (7), we have to prove that,

$$(8)\quad \int_0^\infty F_2(x/t_1)[f_1(t_1) - g_1(t_1)]dt_1 - \int_0^\infty \delta(x/t_1)g_1(t_1)dt_1 \leq 0$$

Since $\delta(x/t_1)$ and $g_1(t_1)$ are nonnegative, it is enough to show that:

$$(9)\quad \int_0^\infty F_2(x/t_1)[f_1(t_1) - g_1(t_1)]dt_1 \leq 0$$

Integrating (9) by parts, we get:

$$(10)\quad \int_0^\infty F_2(x/t_1)[f_1(t_1) - g_1(t_1)]dt_1 = \left\{F_2(x/t_1)(F_1(t_1) - G_1(t_1))\right\}_0^\infty - \int_0^\infty [F_1(t_1) - G_1(t_1)]\frac{\partial F_2(x/t_1)}{\partial t_1}dt_1 \leq 0$$

Recalling that $F_1(\infty) = G_1(\infty) = 1$ and $G_1(0) = F_1(0) = 0$, the first term in the right side of (10) vanishes, and what is left to show is that the second term is nonpositive, i.e., that (11) holds

$$(11)\quad -\int_0^\infty [F_1(t_1) - G_1(t_1)]\frac{\partial F_2(x/t_1)}{\partial t_1}dt_1 = \int_0^\infty \left\{[F_1(t_1) - G_1(t_1)]f_2(x/t_1)x/t_1^2\right\}dt_1 \leq 0$$

The last term is nonpositive since, $f_2(x/t_1) \geq 0$, $x/t_1^2 \geq 0$, and by equation (3), $[F_1(t_1) - G_1(t_1)] \leq 0$.

Notice, also, that $\delta(x_0/t_{1_0}) > 0$ for some value (x_0/t_{1_0}), and hence there is a strict inequality in (7) at least for one value. Thus, $F(x) \leq G(x)$ and $F(x_0) < G(x_0)$ for some value x_0 which implies that the two-period distribution F dominates the two-period distribution G.

THEOREM 2: *Let F^n and G^n be the cumulative distributions of two n-period risks (note that the superscript respresents the number of periods which the cumulative distributions represent). A sufficient condition for F^nDG^n for every nondecreasing utility function is $F_i(x_i) \leq G_i(x_i)$ for all values of x_i, and $F_i(x_{i_0}) < G_i(x_{i_0})$ for some value x_{i_0}, for $i = 1, 2, \ldots, n$*

In fact, only one strict inequality is sufficient to prove this theorem. However, to simplify, we shall not emphasize throughout the paper the need for a strong inequality.

PROOF:

The proof is by induction: For $n = 2$, $F_i(x_i) \leq G_i(x_i)$ $(i = 1, 2)$ implies that F^2 dominates G^2 (Theorem 1). Let us assume that Theorem 2 holds for $n-1$ periods, and prove that it also holds for n periods, i.e., we assume that:

$$F_i(x_i) \leq G_i(x_i)$$

$$(i = 1, 2, \ldots, n-1) \Rightarrow F^{n-1} \text{ dominates } G^{n-1}$$

We now have to show that $F_i(x_i) \leq G_i(x_i)$ $(i = 1, 2, \ldots, n)$ implies that F^n dominates G^n. Note that $F_i(x_i) \leq G_i(x_i)$ (for $i = 1, \ldots, n-1$) implies that F^{n-1} dominates G^{n-1} (by assumption) i.e., $F^{n-1}(\prod_{i=1}^{n-1} x_i) \leq G^{n-1}(\prod_{i=1}^{n-1} x_i)$ for all values $\prod_{i=1}^{n-1} x_i$. Since x_n and $\prod_{i=1}^{n-1} x_i$ are independent, and it is given that $F^{n-1} \leq G^{n-1}$ and $F_n \leq G_n$, one can again use Theorem 1 to prove that $F^n(\prod_{i=1}^n x_i) \leq G^n(\prod_{i=1}^n x_i)$ i.e., F^n dominates G^n.[9]

COROLLARY 1: *If, in addition to the assumption of independence, we also assume stationarity over time of the distributions under consideration, one may conclude that for any two horizons, n_1 and n_2 $(n_2 > n_1)$, the number of options in the efficient set which is appropriate to horizon n_2 is not larger than the number of options in the efficient set which is appropriate to horizon n_1, where the efficient sets are defined for all nondecreasing utility*

[9] Note that in Theorem 1 we did not assume stationarity over time of the distributions. The comparison of the pair (F^{n-1}, F_n) with the pair (G^{n-1}, G_n) does not violate the conditions under which Theorem 1 holds.

functions. This conclusion stems directly from Theorem 1.

Suppose we have two options, F and G. The stationarity assumption implies that, $F_1=F_2 \ldots =F_n$ and $G_1=G_2= \ldots G_n$. For the sake of simplicity, suppose that $n_1=1$ and $n_2=n$. If F_1DG_1, then, by the stationarity assumption, F_i also dominates G_i for all other periods $i=2, \ldots, n$.[10] Hence, by Theorem 1, F^n dominates G^n.

To sum up, an option which is eliminated from the one-period efficient set, is also eliminated from the n-period efficient set. Hence, the number of options in the efficient set of the long horizon cannot exceed the number of options in the efficient set of the short horizon.

It is interesting to note that one can find numerical examples for which the number of elements in the efficient set is strictly decreasing when the horizon increases. Consider the following two-period example, where $F_1=F_2$ and $G_1=G_2$.

Example 1

G_1 (or G_2)		F_1 (or F_2)	
Outcome	Probability	Outcome	Probability
1	1/4	2	1/2
4	3/4	10	1/2

Assuming independence over time, the two-period distributions F^2 and G^2 are:

G^2		F^2	
Outcome	Probability	Outcome	Probability
1	1/16	4	1/4
4	6/16	20	1/2
16	9/16	100	1/4

Since F_1 and G_1 (or F_2 and G_2) intersect, both of them are included in the one-period efficient set. But $F^2(x) \leq G^2(x)$ for every x (with a strict inequality for some values) and, hence, F^2DG^2. This numerical example shows that portfolios which are not eliminated from the one-period efficient set might be eliminated from the multi-period efficient set.[11]

II. Multi-period Dominance in the Case of Risk Aversion

It has been shown that for the one-period investment case, a necessary and sufficient condition for a risk with cumulative probability distribution function F_1 to dominate one with $G_1(F_1DG_1)$ as its cumulative distribution function, for all nondecreasing *concave* utility functions, is:

$$\int_{-\infty}^{x} [G_1(t) - F_1(t)]dt \geq 0$$

for every x, and $G_1 \neq F_1$ for some x_0, where the subscript 1 indicates the distribution for period one. (See articles by Hadar and Russell, Hanoch and Levy, and Rothschild and Stiglitz.) We shall now prove that if F_1DG_1 and F_2DG_2, then, F dominates G, where F and G are the two-period distributions

Before proceeding to state and prove a theorem for the multiperiod risk aversion case we need the following:

LEMMA: *Let $T(x, t_1)=\int_0^x F(t/t_1)dt$, where F is some (cumulative) distribution function. Then,*

$$\frac{\partial T(x, t_1)}{\partial t_1} = T'(x, t_1) \leq 0 \quad \textit{and}$$

$$\frac{\partial^2 T(x, t_1)}{(\partial t_1)^2} = T''(x, t_1) \geq 0$$

PROOF:

$$(12) \qquad T'(x, t_1) = \frac{\partial \int_0^x F(t/t_1)dt}{\partial t_1} = \int_0^x \left(\frac{\partial F(t/t_1)}{\partial t_1}\right) dt$$

Let $f(t/t_1)$ be the density function of $F(t/t_1)$. Then,

[10] The stationarity assumption implies that the number of elements in the efficient sets constructed for each single period (e.g., year) is the same.

[11] The numerical example confirms that the conditions for dominance given in Theorem 1 are sufficient, but not necessary.

$$T'(x, t_1) = -\int_0^x [f(t/t_1)t/t_1^2]dt \leq 0$$

since the integrand is always nonnegative.

The proof that $T''(x, t_1) \geq 0$ is somewhat more involved:

$$T''(x, t_1) = -\frac{\partial \int_0^x [f(t/t_1)t/t_1^2]dt}{\partial t_1}$$

$$= -\int_0^x \frac{\partial[f(t/t_1)t/t_1^2]}{\partial t_1} dt$$

Hence,

$$\text{(13)} \quad T''(x, t_1) = -\left[\int_0^x f'(t/t_1)(-t/t_1^2)(t/t_1^2)dt + \int_0^x f(t/t_1)(-2t/t_1^3)dt\right]$$

Let us expand the first term on the right side of (13):

$$\int_0^x f'(t/t_1)\cdot(t^2/t_1^4)dt = \int_0^x [f'(t/t_1)1/t_1\cdot t^2/t_1^3]dt$$

Integrating by parts (with respect to x), we get:

$$\text{(14)} \quad \int_0^x [f'(t/t_1)(1/t_1)\cdot t^2/t_1^3]dt$$

$$= \{[f(t/t_1)t^2/t_1^3]\}_0^x - \int_0^x f(t/t_1)2t/t_1^3 dt$$

$$= f(x/t_1)x^2/t_1^3 - \int_0^x f(t/t_1)2t/t_1^3 dt$$

Substituting these results in (13), we get:

$$\text{(15)} \quad T''(x, t_1) = f(x/t_1)x^2/t_1^3 \geq 0$$

We can now prove the following theorem:

THEROEM 3: *Let F and G be the cumulative distributions for the two-period horizon of two alternative risks. A sufficient condition for FDG, for every nondecreasing concave utility function* ($\phi'(x) \geq 0$, $\phi''(x) \leq 0$) is:

$$\text{(16)} \quad \int_0^{x_i} [G_i(t_i) - F_i(t_i)]dt_i \geq 0 \quad (i = 1, 2)$$

for every value x_i, with a strict inequality for at least one value.

PROOF:

We have to prove that (17) holds:

$$\text{(17)} \quad \int_0^x [G(t) - F(t)]dt \geq 0$$

(with a strict inequality for some value x_0). Using (5) and (6), equation (17) can be rewritten:

$$\text{(18)} \quad \int_0^x \int_0^\infty G_2(t/t_1)g_1(t_1)dt_1 dt \geq \int_0^x \int_0^\infty F_2(t/t_1)f_1(t_1)dt_1 dt$$

By condition (16), for the second-period distribution we have:

$$\text{(19)} \quad \int_0^x G_2(t/t_1)dt = \int_0^x F_2(t/t_1)dt + \delta(x/t_1)$$

where $\delta(x/t_1) \geq 0$ and $\delta(x_0/t_{1_0}) > 0$ for at least one value (x_0/t_{1_0}). Substituting (19) in (18), we obtain:

$$\text{(20)} \quad \int_0^\infty ([g_1(t_1) - f_1(t_1)] \int_0^x F_2(t/t_1)dt)dt_1 + \int_0^\infty g_1(t_1)\delta(x/t_1)dt_1 \geq 0$$

Since the second term is nonnegative, it will suffice to prove that the first term in (20) is nonnegative.

Integrating the first term in (20) by parts, with respect to t_1, we obtain:

$$\text{(21)} \quad \left\{[G_1(t_1) - F_1(t_1)] \int_0^x F_2(t/t_1)dt\right\}_0^\infty$$

$$-\int_0^\infty [(G_1(t_1)-F_1(t_1)] \frac{\partial \int_0^x F_2(t/t_1)dt}{\partial t_1} dt_1 \geq 0$$

Since the first term equals zero, and $\int_0^{x_1} F_2(t/t_2)dt \equiv T(x, t_1)$ (by definition), it remains to prove that the following term is nonnegative.

$$(22) \quad \int_0^\infty \{[G_1(t_1) - F_1(t_1)]\} \left(-\frac{\partial T(x, t_1)}{\partial t_1}\right) dt_1 \geq 0$$

Note that $\int_0^{x_1}[G_1(t_1)-F_1(t_1)]dt_1 \geq 0$ for every x_1 (by (16)), and $-T(x, t_1)$ is nondecreasing concave (utility) function in t_1 (by the Lemma). Applying (2′) to this case we get,

$$(23) \quad \int_0^{x_1} [G_1(t_1) - F_1(t_1)]dt_1 \geq 0$$

$$\Rightarrow \int_0^\infty [G_1(t_1) - F_1(t_1)]\left(-\frac{\partial T(x, t_1)}{\partial t_1}\right)dt_1 \geq 0$$

Hence inequality (23) holds.

We also have in (20) at least one strict inequality, since for some value x_0/t_{1_0}, $\delta(x_0/t_{1_0}) > 0$. Hence, FDG.

THEOREM 4: *Let F^n and G^n be the cumulative distributions of two n-period risks. A sufficient condition for F^nDG^n for all nondecreasing concave utility functions is $\int_0^{x_i}[G_i(t_i) - F_i(t_i)]dt_i \geq 0$ for every x_i (with a strict inequality for some x_{i_0}), where $i=1, 2, \ldots, n$.*

PROOF:

The proof is by induction and similar to that given in Theorem 2. For $n=2$, it has been proved in Theorem 3 that $\int_0^{x_i}[G_i(t_i) - F_i(t_i)]dt_i \geq 0$ (for every x_i where $i=1, 2$) implies that F^2 dominates G^2, for all risk averters. Assume, that the statement of the theorem holds for $n-1$ period, i.e., $\int_0^{x_i}[G_i(t_i) - F_i(t_i)]dt_i \geq 0$ (for every x_i where $i=1, 2, \ldots, n-1$) implies that F^{n-1} dominates G^{n-1}. We have to prove that the theorem also holds for n-periods. This is straightforward; since F^{n-1} dominates G^{n-1}, one can conclude that

$$\int_{-\infty}^{\prod_{i=1}^{n-1} x_i} [G^{n-1}(t) - F^{n-1}(t)]dt \geq 0 \quad \text{for every } \prod_{i=1}^{n-1} x_i$$

Since by assumption of the theorem,

$$\int_{-\infty}^{x_n} [G_n(t) - F_n(t)]dt \geq 0$$

for every x_n, we may again employ Theorem 3[12] and prove that

$$\int_{-\infty}^{\prod_{i=1}^{n} x_i} [G^n(t) - F^n(t)]dt \geq 0$$

for all values $\prod_{i=1}^n x_i$, i.e., F^n dominates G^n for all risk averters.

COROLLARY 2 *If we are prepared to assume stationarity over time, in addition to independence of the returns, i.e., $F_1=F_2= \ldots =F_n$ and $G_1=G_2= \ldots =G_n$, we may conclude that the number of elements in the efficient set which is constructed for all risk averters does not increase when the investment horizon increases. This conclusion follows from Theorem 3.*

Suppose that in period 1, F_1 dominates G_1, i.e., $\int_0^{x_1}[G_1(t)-F_1(t)]dt \geq 0$ for all values x_1, then, by the stationarity assumption, F_i dominates G_i for all other periods i ($i=1, 2, \ldots, n$), and by Theorem 3, F^n dominates G^n. In other words, any option which is eliminated from the risk averters' one-period efficient set is also eliminated from the n-period efficient set. The efficient set in each single period is identical, and the n-period efficient set is no greater than the single period efficient set. Again, as in the previous case, we can find examples where the efficient set is strictly decreasing when

[12] Note again that, by assumption, x_n and $\prod_{i=1}^{n-1} x_i$ are independent, i.e., by looking at the results of the first $n-1$ periods one cannot improve prediction with respect to the nth period.

the investment horizon increases.[13] We shall again use Example 1 above to illustrate this argument: F_1 does not dominate G_1 (i.e., $F_1 D G_1$) for all risk averters since,

$$\int_0^4 [G_1(t_1) - F_1(t_1)]dt_1 = -1/4 < 0$$

But, also, G_1 does not dominate F_1

$$\int_0^2 [F_1(t_1) - G_1(t_1)]dt_1 = -1/4 < 0$$

Hence, F_1 and G_1 are included in the one-period efficient set. Assuming stationarity over time $F_1 = F_2$, $G_1 = G_2$, and hence also $F_2 D G_2$, $G_2 D F_2$, which means that F_2 and G_2 are included in the second period efficient set.

Looking back at Example 1, we can see that, in the two-period horizon, $F^2(x) \leq G^2(x)$, for every x and for some value x a strict inequality holds. Consequently $\int_0^x [G^2(t) - F^2(t)dt] \geq 0$ (with a strict inequality for some value x_0) and, hence, F^2 dominates G^2. Therefore, in this example the risk averters' efficient set decreases when the investment horizon increases.

IV. Concluding Remarks

This study investigated the role of the investment horizon in determining the efficient set of portfolios. We found that, unlike the one-period case, the efficiency criteria for the multi-period case are sufficient but not necessary for dominance. If a portfolio is eliminated from each single period efficient set it will also be eliminated from the multi-period efficient set, but the inclusion of a portfolio in *each* single period efficient set does not guarantee that this portfolio will be included in the multi-period efficient set.

Thus, an investment consultant who screens all the available portfolios should construct the efficient set for each group of investors according to their investment horizon. However, the consultant may construct the efficient set for the shortest relevant horizon, knowing that the efficient set which is appropriate for longer horizons must be a subset of (or identical to) the efficient set which has been obtained for the short horizon. Nevertheless, one should keep in mind that the suggested procedure is valid only under the case in which probability functions are stationary over time. Otherwise, the consultant can not limit his attention only to the first period distributions.

The magnitude of the decrease in the efficient set resulting from an increase in the investment horizon is left to be investigated either empirically or by simulation.

[13] This means that the longer the horizon the smaller the efficient set. Similar results have been obtained by Nils Hakansson, who assumes that the objective function of the investors is to maximize the geometric average return over time. In Hakansson's analysis, when $n \to \infty$, the efficient set includes only one portfolio, which is also the optimal one. It is of interest to note that Menachem Yaari, who analyzes the individual's multi-period investment-consumption decision, obtains similar results: the longer the horizon, the more certain the optimal consumption and when $n \to \infty$, the individual investment-consumption decisions are determined as with certainty.

REFERENCES

K. Arrow, *Aspects of the Theory of Risk Bearing*, Helsinki 1965.

P. H. Cootner, *The Random Character of Stock Market Prices*, rev. ed., Cambridge, Mass. 1967.

J. Hadar and W. R. Russell, "Rules for Ordering Uncertain Prospects," *Amer. Econ. Rev.*, Mar. 1969, *59*, 25–34.

N. H. Hakansson, "Multi-Period Mean-Variance Analysis: Toward a General Theory of Portfolio Choice," *J. Finance*, Sept. 1971, *26*, 857–84.

G. Hanoch and H. Levy, "The Efficiency Analysis of Choices Involving Risk." *Rev. Econ. Stud.*, July 1969, *36*, 335–46.

H. Levy and M. Sarnat, *Investment and Portfolio Analysis*, New York 1972.

——— and ———, "International Diversification of Investment Portfolios," *Amer. Econ. Rev.*, Sept. 1970, *60*, 668–75.

J. Lintner, "Security Price, Risk, and Maximal Gains from Diversification," *J. Finance*, Dec. 1965, *20*, 587–615.

H. M. Markowitz, *Portfolio Selection*, New York 1959.

J. Mossin, "Optimal Multi-Period Portfolio Policies," *J. Bus., Univ. Chicago,* Apr. 1968, *41,* 215–29.

J. Pratt, "Risk Aversion in the Small and in the Large," *Econometrica,* Jan. 1964, *32,* 122–36.

J. P. Quirk and R. Saposnik, "Admissibility and Measurable Utility Functions," *Rev. Econ. Stud.,* Feb. 1962, *29,* 140–46.

M. Rothschild and J. E. Stiglitz, "Increasing Risk: I. A Definition," *J. Econ. Theor.,* Sept. 1970, *2,* 225–43.

W. F. Sharpe, "Capital Assets Prices: A Theory of Market Equilibrium Under Conditions of Risk," *J. Finance,* Sept. 1964, *19,* 425–42.

J. Tobin, "The Theory of Portfolio Selection," in F. H. Hahn and F. P. R. Brechlings, eds., *The Theory of Interest Rates,* London 1965.

M. E. Yaari, "A Law for Large Numbers in the Theory of Consumer's Choice Under Uncertainty," working pap. no. CP-330, Berkeley Center Res. Manage. Sci., Univ. California, Mar. 1971.

Spot Speculation, Forward Speculation, and Arbitrage: Comment

By HOUSTON H. STOKES*

In his famous article in *Staff Papers*, S. C. Tsiang showed that the return to spot speculation was equal to the return to forward speculation plus the return to interest arbitrage. This note argues that this result is only strictly true in all cases if there is no government (central bank) intervention in the forward market. In some situations of government intervention, Tsiang's result will not hold. As a consequence there is a loss of generality of Tsiang's analysis unless other conditions are specified. In this paper I develop the conditions to tell whether there is government intervention and if such intervention is sufficient to invalidate Tsiang's analysis.

I. The Model

If F is the 90-day forward rate,[1] S is the spot exchange rate, S^e the expected spot rate in 90 days, and i^f and i^d the foreign and domestic 90-day interest rates, respectively, then interest parity[2] implies

$$F(1 + i^f) = S(1 + i^d) \tag{1}$$

If

$$F^* \equiv \frac{S(1 + i^d)}{(1 + i^f)} \tag{2}$$

there will be an outflow (inflow) of funds from the domestic country if F is greater (less) than F^*. If C_{at} is defined as the desired stock of forward contracts held by arbitragers in period t, then[3]

$$C_{at} = \phi(F - F^*) \quad \text{where } \phi < 0 \tag{3}$$

Forward speculators will sell (buy) the foreign currency forward if F is greater (less) than S^e. If C_{st} is the desired stock of forward contracts held by forward speculators in period t then

$$C_{st} = \Psi(F - S^e) \quad \text{where } \Psi < 0 \tag{4}$$

Parity in the spot speculation market occurs when

$$S(1 + i^d) = S^e(1 + i^f) \tag{5}$$

From equations (5) and (2) we note that spot speculators will acquire the foreign (domestic) currency spot if S^e is greater (less) than F^*. If $C_{\hat{s}t}$ is the desired stock of the domestic currency held by spot speculators in period t then

$$C_{\hat{s}t} = Z(S^e - F^*) \quad \text{where } Z < 0 \tag{6}$$

We can plot equations (3) and (4) as the AA and SS curves where the intersection of SS and AA with the vertical axis is S^e and F^*, respectively. This has been done in Figures 1 and 2.[4] In the absence of governmental intervention $F^* \geq F \geq S^e$ (Figure 1) or $S^e \geq F \geq F^*$ (Figure 2). Since the return to

* Assistant professor of economics, University of Illinois at Chicago Circle. I am indebted to Robert Mundell, Harry Johnson, and other members of the University of Chicago International Trade Workshop. Zoran Hodjera of the IMF made many helpful suggestions on an earlier draft. An anonymous referee also contributed. Any remaining errors are my responsibility.

[1] All exchange rates are quoted as the domestic currency price of one unit of the foreign currency. Time subscripts are ignored.

[2] This interest parity condition, the usual one, is a simplified version of the one Tsiang uses. If one thinks in Tsiang's terms one can "adjust" the interest rates for the subjective marginal convenience yield. This simplification does not affect the analysis.

[3] If C_{at} is negative (positive) arbitragers have made an outflow (inflow).

[4] The graphical treatment of the model is based on B. Reading's work. This model is a simplified version of Tsiang's; I have used it in my 1972 article. If $\phi = -\infty$ then there is perfect capital mobility and the AA curve is perfectly elastic. Here arbitragers determine the forward rate. If $\Psi = -\infty$ then the SS curve is perfectly elastic and speculators determine the forward exchange rate. It is impossible for $\phi = -\infty$ and $\Psi = -\infty$ at the same time except in the trivial case where $S^e = F^*$ or the SS curve lies on top of the AA curve.

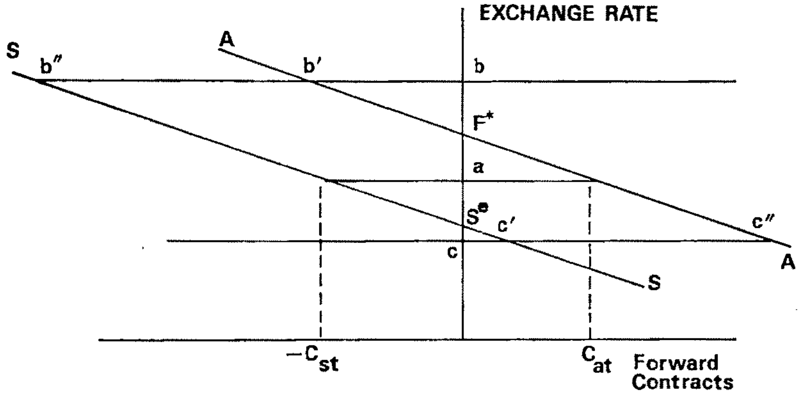


FIGURE 1. NOTE: a = NONINTERVENTION LEVEL OF F

forward speculation is ($|S^e-F|$), the return to interest arbitrage is ($|F^*-F|$) and the return to spot speculation is ($|S^e-F^*|$) it clearly follows that in the absence of governmental intervention in the forward exchange market, Tsiang is correct; the return to spot speculation is equal to the return to forward speculation plus the return to interest arbitrage. This reduces to

$$|S^e - F^*| = |S^e - F| + |F - F^*| \tag{7}$$

We note that in this case

$$|C_{st}| = |C_{at}| \tag{8}$$

If the government intervenes in the forward market it is possible to have[5] any of the following cases.

[5] In my 1972 paper, I developed conditions concerning the effectiveness of forward intervention. By differentiating (3) and (6) with respect to F we can derive an expression for the net flow of money to the domestic country (dM_d) due to forward intervention. This reduces to

$$dM_d = dF\left[\frac{\partial S^e}{\partial F}Z + \phi\right] \quad \text{where} \quad \phi \neq -\infty$$

and

$$dM_d = dI\left[\frac{\partial S^e}{\partial F}Z + 1\right] \quad \text{where} \quad \phi = -\infty$$

and I = the dollar amount of forward contracts of the foreign currency sold by the central bank. This analysis builds on the effect of forward intervention on the position of the AA and SS schedules mentioned briefly by Tsiang but not worked out.

$$F \geq F^* \geq S^e \tag{9}$$

$$F^* \geq S^e \geq F \tag{10}$$

$$F \geq S^e \geq F^* \tag{11}$$

$$S^e \geq F^* \geq F \tag{12}$$

In all these cases, Tsiang's result does not hold. Equation (9) represents the government bidding the forward rate up to b starting from an initial situation represented by point a in Figure 1. The government has bought $bb'+bb''$ forward contracts from the arbitragers and forward speculators, respectively. Since F is now no longer between F^* and S^e, Tsiang's statement that the return to spot speculation $|S^e-F^*|$ is equal to the return to forward speculation $|S^e-F|$ plus the return to arbitrage $|F-F^*|$ is not correct. In this situation the return to forward speculation $|F-S^e|$ is equal to the return on arbitrage $|F-F^*|$ plus the return to spot speculation $|S^e-F^*|$.

In equation (10) the government starting from initial equilibrium in Figure 1 has sold contracts to the arbitragers and forward speculators driving the forward rate to c. Total government sales of forward contracts are equal to $cc'+cc''$. Here again Tsiang's proof does not hold. In this situation the return to arbitrage $|F-F^*|$ is equal to the return to spot speculation $|S^e-F^*|$ plus the return to forward speculation $|S^e-F|$.

In equation (11) the government starting

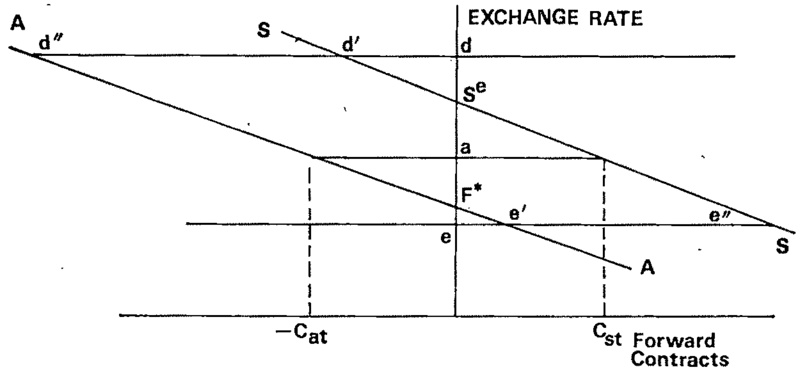


FIGURE 2. NOTE: a=NONINTERVENTION LEVEL OF F

from an initial equilibrium at point a in Figure 2 has bid the forward rate up to d by buying dd' from the forward speculators and dd'' from the arbitragers. Tsiang's proof no longer holds. We find that the return to arbitrage $|F-F^*|$ is equal to the return to spot speculation $|S^e-F^*|$ plus the return to forward speculation $|S^e-F|$.

In equation (12) the government has intervened to disturb the initial equilibrium represented by Figure 2 to push the forward rate down to e by selling ee' to the arbitragers and ee'' to the forward speculators. Here the return to forward speculation $|S^e-F|$ is equal to the return to spot speculation $|S^e-F^*|$ plus the return to arbitrage $|F-F^*|$.

Summarizing: In equations (9) and (12) instead of Tsiang's result, we find that the return to forward speculation is equal to the return to interest arbitrage plus the return to spot speculation. In equations (10) and (11), the return to interest arbitrage is equal to the return to spot speculation plus the return to forward speculation. These results are not surprising since if equation (9) represents the domestic country, equation (12) looks at the same situation from the point of view of the foreign country. Equations (10) and (11) are similarly related.

In short, only when

$$F^* \geq F \geq S^e \tag{13}$$

or

$$S^e \geq F \geq F^* \tag{14}$$

does Tsiang's result hold.

Equation (13) or (14) will hold if

$$|C_{at}| = |C_{st}| \tag{8}$$

and

$$C_{at} \neq C_{st} \tag{15}$$

If equation (8) and (15) do not hold we know there is government intervention.

Equation (13) and (14) can hold if $|C_{at}| \neq |C_{st}|$ if the government intervention has not driven the forward rate outside the band of spot speculation profitability, i.e., if C_{st} and C_{at} are not the same sign.[6]

The above findings are very important since Tsiang notes:

> . . . a speculator who speculates on the spot exchange market may, in fact, be regarded as acting implicitly in the combined capacity of an interest arbitrager and a forward exchange speculator . . . the tentative assumption stated above that all speculation in foreign exchange is carried on in the forward market, does not impair the generality of our analysis as long as speculators who in effect speculate in the spot market are treated according to their dual capacity, namely, first as interest arbitragers and then as speculators in forward exchange. [p. 92]

[6] Equation (15) is necessary if we are to take into consideration the possibility of an intersection of the AA and SS curves. In such a situation intervention could drive the forward rate to a position where equation (8) holds but where neither equation (13) or (14) is satisfied.

My paper has questioned Tsiang's basic assumption. I have shown that although Tsiang's analysis holds in all cases of nonintervention it holds with intervention only if equation (13) or (14) are met. If intervention causes equations (9), (10), (11), or (12) to hold, Tsiang's results lose generality. The most likely possibility would be a situation such as equation (12) where the government in response to a rapidly rising expected spot rate is pushing the forward rate down. This appeared to be the case of the United Kingdom in 1967 where during the crisis prior to the devaluation, the interest arbitragers were coming to the United Kingdom while the speculators were leaving. In short, this analysis suggests that for empirical work Tsiang's model must be used with caution since in the last decade prolonged periods of nonintervention have been the exception rather than the rule.

REFERENCES

B. Reading, "The Forward Pound, 1951–1959," *Econ. J.*, June 1960, *70*, 304–19.

H. H. Stokes, "The Effect of Forward Exchange Intervention: Comment," *Southern Econ. J.*, Oct. 1972, *39*, 311–16.

S. C. Tsiang, "The Theory of Forward Exchanges and the Effects of Government Intervention on the Foreign Exchange Market," *Int. Monet. Fund Staff Pap.*, Apr. 1959, *7*, 75–106.

Spot Speculation, Forward Speculation, and Arbitrage: A Clarification and Reply

By S. C. TSIANG*

Houston Stokes' comment on my paper seems to be based on a misinterpretation of a passage of my 1959 article, which perhaps calls for further clarification. He asserts that I said somewhere in my paper that " . . . the return to spot speculation was equal to the return to forward speculation plus the return to interest arbitrage," (p. 996) but he does not bother to point out where I am supposed to have said this.

Searching through that article, which I wrote fourteen years ago, however, I could not find any sentence to that effect. What I said there in this connection was the following:

> Thus a speculator who speculates on the spot exchange market may, in fact, be regarded as acting implicitly in the combined capacity of an interest arbitrager and a forward exchange speculator. If he speculates by buying spot foreign exchange (instead of buying forward exchange), what he is doing is in fact equivalent to acting first as an interest arbitrager in buying spot exchange against a sale of an equal amount of forward exchange, and then as a forward speculator in buying the forward exchange from himself (in the former capacity) in the expectation that the future spot rate will rise higher than the current forward rate and yield him a pure speculative gain. If he speculates by selling spot exchange (instead of forward exchange), he is in fact acting first as an interest arbitrager in selling spot exchange against an equal amount of forward purchase, and then as a forward speculator in selling the forward exchange to himself in the former capacity in the expectation that the future spot rate will turn out to be lower than the current forward rate. [p. 92]

This statement is tautologically always true, even though the statement which Stokes attributed to me may not always be true.

There are no difficulties in the first two cases which he depicted in Figure 1 and Figure 2, i.e., $F^*>F>S^e$ and $S^e>F>F^*$, where F is forward exchange rate, $F^*=S(1+i^d)/(1+i^f)$ is the forward rate that would correspond to the interest parity (S being defined as the current spot rate and i^d and i^f the domestic and foreign interest rates, respectively), and S^e is the speculators' expected future spot rate. Stokes thinks that in the absence of government intervention, these are the only cases possible. This is open to question, but I shall postpone its discussion until later. In four other possible cases, he claims that I was wrong. For instance let us take the case $F>F^*>S^e$, i.e., his inequality (9).

In this case arbitragers would presumably hold spot foreign exchange and cover the holdings by selling equal amount forward. The profit of arbitrage per unit of foreign exchange would be $F-F^*$ if we neglect the convenience yields of spot funds on hand. The spot speculator would go short on spot exchange and expect to make the profit of F^*-S^e per unit of foreign exchange. The forward speculator would go short on forward exchange and expect to make the profit of $F-S^e$ per unit. Thus obviously the return to spot speculation is not equal to the return to forward speculation plus that to covered interest arbitrage. Rather it is the difference between the return to forward speculation and that to covered interest arbitrage. Therefore, he claims that I am wrong.

However, if he had read me more carefully, he should have seen that what is said in the quotation above is literally correct even in this anomalous case. For what I said was that " . . . if he (the speculator) specu-

* Cornell University.

lates by selling spot exchange, he is in fact acting first as an interest arbitrager in *selling* spot exchange against an equal amount of forward *purchase*, and then as a forward speculator in selling the forward exchange to himself . . . " It is clear from this sentence, that speculative *selling* of spot foreign exchange implies an arbitrage operation of *selling* spot covered by *buying* forward and then a speculative sale of forward exchange. In this anomalous case, however, the implied arbitrage operation in the direction I specified would incur a loss of (F^*-F). So the spot speculator's expected return from his spot selling is equal to the expected return to forward speculative selling minus the loss from the implied arbitrage operation. The arithmetic would then come out exactly right. Stokes' difficulty arises because he misinterprets my statement to mean that a spot speculator's expected return is equal to the expected return to forward speculation plus the return to interest arbitrage in whichever direction that is profitable at the time. Had he read my paper more carefully, he would have noticed that the arbitrage operation and accompanying forward speculation implied in a spot speculative sale (or purchase) should be in the specific directions clearly specified in the above paragraph, not just any pair of such operations in the directions that is profitable under the circumstances.

All the four cases in which he claims that he has caught me wrong can be explained in this way.

Stokes might perhaps shift his ground of criticism and go on to protest that if I claim my theory to be general, how can I assume that the spot speculators would implicitly do obviously irrational things, such as arbitraging in a direction opposite to what the interest parity and forward premium would indicate, or selling forward exchange for speculation when his own expected future spot rate is higher than the forward rate. The answer to this paradoxical question is that in all those four anomalous cases, it is in fact irrational for any one to undertake spot speculation. The fact that when spot speculation is decomposed into two implicit actions, one of these would obviously be irrational, merely reflects the inherent irrationality of such activity in those cases. For instance, in the case of his (9), $F>F^*>S^e$, the speculator who sells spot is obviously acting irrationally if he has equal access to the forward market as I assumed. For why should any one incur additional net interest cost in selling spot forward exchange in order to speculate, if he can sell forward exchange and expect to make a greater profit while running exactly the same risk arising from the uncertainty of expected future rate.

Readers with patience can easily reason out why in the other three anomalous cases also, spot speculation would itself be an irrational behavior. To decompose spot speculation into two implicit component acts, one of which is in the irrational direction, does not mean that we are imputing some irrationality upon a rational behavior. It merely brings out in stronger relief the inherent irrationality of spot speculation under those circumstances.

The reason why I wanted to subsume spot speculation under forward speculation and covered interest arbitrage is to simplify the theory of the joint determination of forward and spot exchange rates such that these rates can be said to be determined by actions of only three groups of participants: forward speculators, arbitragers, and commercial coverers, apart from government authorities who might intervene from time to time. Reviewing this theory after the lapse of fourteen years, it is gratifying to find that this theory is indeed quite general as I claimed then. For in the two normal cases, spot speculation can certainly be split into two implicit activities, covered arbitrage and forward speculation, each following its own rational principles of behavior. In the four anomalous cases that Stokes listed, spot speculation is an irrational behavior and can therefore be ruled out.

Having answered Stokes' criticism, it is my turn to offer him some comments in return. Stokes thinks that in the absence of governmental intervention the only cases possible are the two normal ones, i.e., $F^*>F>S^e$ and $S^e>F>F^*$. No explicit reason is given for

this, but one might perhaps guess that the reason at the back of his mind is that he somehow regards that without governmental intervention his equation (8) would hold, i.e.,

$$|C_{st}| = |C_{at}|$$

by which he probably meant that the desired stock of forward contracts held by forward speculators must equal the desired stock of forward contracts held by arbitragers in the opposite direction, i.e., $C_{st} = -C_{at}$.

Why should this be true is a mystery to me. Certainly my theory would not lead to such a conclusion at all. Stokes apparently thought that, without government intervention, forward speculators and arbitragers would be the only groups of operators on the forward exchange market, so that the demand of the one must be matched by the supply of the other.

This is patently not true; for he has left out a most important group of operators on the forward market, viz., traders seeking covering for their foreign exchange commitments. The chief social function of a forward exchange market is essentially to provide hedging or covering to traders. If traders do not deal in forward exchange at all, forward exchange market would lose its main *raison d'être*. When commercial coverers operate in forward market, there is no reason why Stokes' equation (8), or rather $C_{st} = -C_{at}$, should hold at all. I cannot see any reason why Stokes should abandon my diagram 1 on page 96 in favor of his grossly simplified and incomplete diagram to show the forces at play on the forward exchange market and the determination of the forward rate. Had he used my diagram, he would see that it is not impossible to have, say, a situation like (10) $F^* > S^e > F$, without any government intervention at all. As illustrated in Figure 1, there might be an excess supply of forward exchange on the part of commercial coverer (because of an export surplus). The excess supply may be absorbed by both the arbitragers and the speculators who for some reasons or other might expect only a mild appreciation. Thus the export surplus AB, which, under the assumption that traders cover all their trade commitments, would coincide with the excess supply of forward exchange on the part of commercial traders, is divided between forward speculators, who absorb $AC = FD$, and the interest arbitragers, who absorb $CB = EF$.

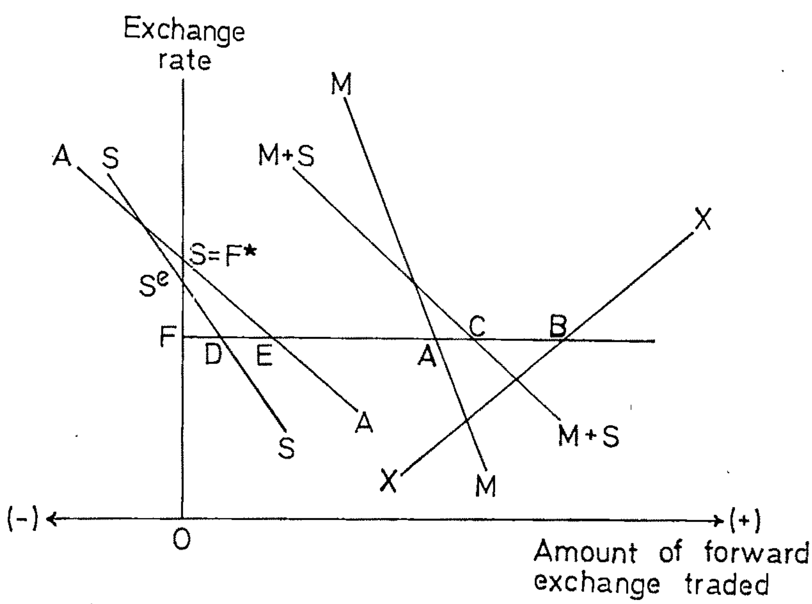


FIGURE 1

In this case, the generality of my theory still stands; viz., that both the forward and the spot rates of exchange can be said to be determined by the actions of three groups of participants: the commercial coverers, the arbitragers, and the forward speculators. For in this case, spot speculation, which would occupy spot funds just as interest arbitrage, but which would yield a return that is not only uncertain and risky but also expected to be smaller than the sure return of interest arbitrage, is certainly irrational and therefore can be ignored.

REFERENCE

S. C. Tsiang, "The Theory of Forward Exchanges and the Effects of Government Intervention on the Foreign Exchange Market," *Int. Monet. Fund Staff Pap.*, Apr. 1959, *7*, 75–106.

Pitfalls in Financial Model Building: Comment

By KEVIN CLINTON*

Mark Ladenson's recent elaboration of the point made by William Brainard and James Tobin, that security demand functions must be consistent with the balance sheet constraint, overcomplicates a simple issue. Furthermore, the two contributions by Ladenson and by Brainard and Tobin, taken together, are incomplete and misleading in certain respects. This note gives a general statement of the conditions for exact exhaustion of the balance sheet constraint which includes the Brainard and Tobin and Ladenson conditions as special cases, using derivations that are much simpler than those in Ladenson's article.

The demand functions with which I am concerned are written as linear functions, homogeneous in net wealth, W. Where y_{it}^* is the desired amount of asset i at time t, these functions are

$$(1) \qquad \overset{*}{y}_{it} = \beta_i W_t + \sum_{k=1}^{K} \beta_{ik} x_{kt} W_t \qquad i = 1, 2, \ldots, n$$

In addition to wealth there are K explanatory variables x_{kt} (interest rates, income, and the like) for the n assets. The β's are relevant coefficients.

Assuming that there is a generalized stock adjustment mechanism at work with parameters $\alpha_{ij}(i, j=1, 2, \ldots, n)$ the asset demand functions become

$$(2) \quad \Delta y_{it} = \sum_j \alpha_{ij}(\overset{*}{y}_{jt} - y_{j,t-1}) + \gamma_i \Delta W_t$$

where Δ is the first difference operator and γ_i is a coefficient which determines the initial allocation of a change in wealth into the component assets. Substitute (1) into (2):

$$\begin{aligned}(3) \quad \Delta y_{it} &= \sum_j \alpha_{ij}\left(\beta_j W_t + \sum_{k=1}^{K} \beta_{jk} x_{kt} W_t - y_{j,t-1}\right) + \gamma_i \Delta W_t \\ &= \sum_j \alpha_{ij}\beta_j W_t + \sum_j \alpha_{ij} \sum_k \beta_{jk} x_{kt} W_t - \sum_j \alpha_{ij} y_{j,t-1} + \gamma_i \Delta W_t\end{aligned}$$

The restrictions on the static model (1) are that $\sum_i \beta_i = 1$ and $\sum_i \beta_{ik} = 0$ for all k. Assume that $\sum_i \alpha_{ij}$ is constant over j. Sum (3) over i taking the foregoing into account:

$$(4) \qquad \sum_i \Delta y_{it} = \sum_i \sum_j \alpha_{ij}\beta_j W_t - \sum_i \sum_j \alpha_{ij} y_{j,t-1} + \sum_i \gamma_i \Delta W_t$$

In view of the fact that $\sum_i y_{it} = W_t$, (4) can be rearranged into

$$(5) \quad W_t = \left(\sum_j \beta_j \sum_i \alpha_{ij} + \sum_i \gamma_i\right) W_t + \left(1 - \sum_i \alpha_{ij} - \sum_i \gamma_i\right) W_{t-1}$$

or

$$(6) \quad W_t = \left(\sum_i \alpha_{ij} + \sum_i \gamma_i\right) W_t + \left(1 - \sum_i \alpha_{ij} - \sum_i \gamma_i\right) W_{t-1}$$

It is therefore sufficient[1] for the equality in (4) to hold that

$$(7) \qquad \sum_i \alpha_{ij} = 1 - \sum_i \gamma_i \qquad \text{for all } j$$

* Department of banking and financial analysis, Bank of Canada, Ottawa.

[1] As Ladenson points out, only sufficiency is at stake in this discussion, despite Brainard's and Tobin's reference to necessity.

Brainard and Tobin set $\sum_i \gamma_i = 1$, so that $\sum_i \alpha_{ij} = 0$. Ladenson's additional restriction is to set $\sum_i \gamma_i = 0$, so that $\sum_i \alpha_{ij} = 1$. These two restrictions omit a number of admissible and interesting specifications. Consider, for example, the following system:[2]

$$(8) \quad \alpha_{ij} = \begin{cases} \alpha = 1 - \sum_i \gamma_i \neq 0, & \text{for } i = j \\ 0, & \text{for } i \neq j \end{cases}$$

This system is consistent with (7). In view of this consistency it is interesting to find that Brainard and Tobin write, "If no cross effects were allowed in the explicit equations of adjustment . . . , then the counterparts of all the own adjustment coefficients specified would be loaded into the implicit adjustment equation for [the asset omitted by Walras' law]".[3] That is, if a lagged dependent variable appears in one equation then it must appear in at least one other equation also to provide for maintenance of the balance sheet identity Well, I have just provided a counterexample: there are no cross effects ($\alpha_{ij} = 0$ for $i \neq j$) and there is no omitted equation.

If you want to avoid the pitfalls into which Brainard and Tobin stumbled, as well as the pitfalls which they found, you should mark them clearly. As a clear and simple warning of pitfalls in dynamic financial models, I offer conditions (7).

[2] A specification actually used in the econometric model *RDX2*.

[3] See p. 106. The phrase in brackets is mine.

REFERENCES

W. C. Brainard and J. Tobin, "Pitfalls in Financial Model Building," *Amer. Econ. Rev. Proc.* May 1968, *58*, 99–122.

J. F. Helliwell et al., *The Structure of DRX2*, Bank of Canada Staff Res. Study no. 7, Ottawa 1971.

M. L. Ladenson, "Pitfalls in Financial Model Building: Some Extensions," *Amer. Econ. Rev.* Mar. 1971, *61*, 179–86.

Pitfalls in Financial Model Building: Reply and Some Further Extensions

By Mark L. Ladenson*

Kevin Clinton has provided an interesting counterexample to the verbal argument that the omission of cross-adjustment coefficients necessarily misspecifies the system of asset adjustment proposed by William Brainard and James Tobin.[1] However, Clinton's specification is not a counterexample to any of the formal propositions developed in my paper. As Clinton observes in a footnote, I dealt with two alternative sets of *sufficient* conditions for consistency of that system. He has presented a third alternative. All three of these alternatives are special cases of the complete set of *necessary and sufficient* conditions for consistency. I did not develop these latter conditions in my paper since, at the time I wrote it, I did not recognize their economic interpretation. This interpretation can now be provided. I therefore welcome the opportunity to develop the complete set of conditions. I will *derive* as one of these conditions the proposition that all columns of the adjustment coefficient matrix, A, have the same sum (i.e., $\sum_i \alpha_{ij}$ is constant for all j). This is a proposition which Clinton needs in order to rearrange his equation (4) into his equation (5), and which he simply *assumes*. More important, however, will be our discussion of the meaning of the conditions.

We begin with equation (11′) of my 1971 paper in this *Review:*

$$(11') \quad \Delta W_t = r'AB_1X_1W_t + r'ABX_tW_t - r'Ay_{t-1} + r'\gamma_{\Delta W}\Delta W_t$$

Since equation (2) of the paper, $r'y_t = W_t$, holds identically in all time periods, and since $W_t \equiv W_{t-1} + \Delta W_t$, it follows that:

$$W_t = r'y_{t-1} + \Delta W_t$$

We substitute this last identity into the first term on the right-hand side of (11′), recall that X_1 is identically equal to unity, and collect terms to get

$$(I - r'\gamma_{\Delta W} - r'AB_1)\Delta W_t - (r'AB)X_tW_t - (r'AB_1r' - r'A)y_{t-1} = 0$$

This last relation will be satisfied if, and only if,

$$(11''\text{a}) \quad (I - r'\gamma_{\Delta W} - r'AB_1) = 0$$

$$(11''\text{b}) \quad r'AB = 0$$

$$(11''\text{c}) \quad r'A(I_5 - B_1r') = 0$$

The set of equations (11″) constitutes the complete set of necessary and sufficient conditions for consistency of the Brainard-Tobin system. Taking these conditions in reverse order, we will discuss their meaning and their implications for Clinton's counterexample. The third of them, (11″c), is indeed the proposition merely assumed by him: all columns of the adjustment coefficient matrix must have the same sum. For Clinton's counterexample to satisfy the condition, the columns of A must all sum to zero or the elements of the vector B_1 must sum to unity. Since Clinton (following the derivation in my original paper) utilized the latter restric-

* Assistant professor of economics, Michigan State University. I am deeply indebted to my colleague Robert H. Rasche for invaluable aid in preparing this note. I must also acknowledge my continuing debt to Walter D. Fisher, although I have not discussed the note with him. I, of course, accept all responsibility for errors.

[1] Clinton's specification that the change in each asset depends only on an own-adjustment coefficient multiplied by the gap between desired and actual values of that asset plus a coefficient multiplied by the total change in wealth has been previously proposed by Gregory Chow. Clinton has shown that the balance sheet restrictions imply that in Chow's specification all own-adjustment coefficients must be equal to each other. It should also be noted that the ordinary least squares estimates of each equation will not satisfy this restriction. One would probably want to use the estimation method devised by Arnold Zellner, though allowance would have to be made for the fact that the errors add to zero across equations for each observation.

tion in the course of his work, the counterexample satisfies this condition.

The elements of the matrix AB are impact multipliers. They give the impact effect on the assets (*ceteris paribus*) of changes in the exogenous variables other than wealth (interest rates and income). Equation (11″b) asserts that each column of this matrix must be equal to zero. That is, the impact effect of any interest rate or income change summed over all assets must be zero. For Clinton's counterexample to satisfy the condition, the columns of A must sum to zero or the columns of the matrix of structural coefficients, B, must sum to zero. Since Clinton (following the derivation in my original paper) utilized the latter restriction in the course of his work, his counterexample satisfies this condition.

Equation (11″a) deals with coefficients which give the effect of a change in net worth on the assets. These are the elements of the vector $\gamma_{\Delta W}$ and of the vector AB_1. It will be recalled that the elements of the former vector are the coefficients of the variable ΔW, and the elements of the latter vector are the coefficients of the variable W. Thus, in reduced form, the latter vector was named γ_1. Since its elements are the product of adjustment coefficients and structural coefficients they are impact multipliers. The influence of changes in wealth on the asset adjustment process is represented by the elements of both vectors. Equation (11″a) states that the sum of these influences must be unity; that the proportions in which a change in wealth is distributed across the portfolio add to 100 percent.

Unfortunately, as I noted in my paper, the complete system of adjustment equations cannot be estimated since an exact linear relation, $W_t = r'y_{t-1} + \Delta W_t$, prevails between the variables for all possible observations. As a result, if we wish to retain the symmetry of the matrix of adjustment coefficients we are not able to identify the elements of the vectors γ_1 and $\gamma_{\Delta W}$. (Equations (13Ra) and (13Rb) of my paper.) I suggested that we could deal with this problem by assuming all the elements of γ_1, on the one hand, or of $\gamma_{\Delta W}$, on the other, are equal to zero.[2] Equation (11″a) tells us that if we utilize the first of these alternatives, the elements of $\gamma_{\Delta W}$ sum to one, and if we utilize the second, the elements of γ_1 sum to one. In the former case we are setting the value of the impact multipliers equal to zero; we are assuming that the complete effect of a change in wealth on assets is represented by the elements of $\gamma_{\Delta W}$. Since these are not multiplied by the adjustment coefficient matrix, this case, as noted in my paper, involves the assumption that the assets adjust instantaneously to a change in wealth. On the other hand, I argued that if we utilize the other alternative, we assume that such adjustment occurs with a distributed lag. This point should now be more apparent, as we now recognize the elements of γ_1 to be impact multipliers.

It is possible to say more about patterns of adjustment and the meaning of these coefficients. In the case in which we set the elements of $\gamma_{\Delta W}$ equal to zero, the entire effect on the assets of a change in wealth is represented by the elements of γ_1. Since these involve multiplying the structural coefficients, the elements of B_1, by the same adjustment coefficients involved in the matrix AB, we are assuming that the assets adjust to a change in wealth at the same speed with which they adjust to a change in one of the exogenous variables (interest rates or income). Equation (11″a) does not, of course, require all the elements of γ_1 or of $\gamma_{\Delta W}$ to be equal to zero. The sums of the elements of both vectors might be positive and, added together, equal to one. For example, in Clinton's specification the elements of γ_1 sum to α and those of $\gamma_{\Delta W}$ to $(1-\alpha)$. In fact, from a purely formal viewpoint, we cannot exclude the possibility that the sum of the elements of γ_1 is positive, that of $\gamma_{\Delta W}$ is *negative*, and the two sums, added together, equal one. What does it mean if the adjustment equation of one of the assets has positive coefficients on both the W and ΔW variables? We have just seen that the co-

[2] At this point in the paper I made an error which is corrected below.

efficient of W involves the same speed of adjustment to a wealth-induced portfolio disequilibrium as to an interest-rate-induced disequilibrium. If the coefficient on ΔW is also positive it means that the effect of a change in wealth on the asset adjustment process is reflected both in the elements of γ_1 (the coefficient of W) and in the elements of $\gamma_{\Delta W}$ (the coefficient of ΔW). Therefore the asset in question adjusts more quickly to a change in net worth than to a change in interest rates. In parallel fashion, if the coefficient on ΔW is negative it means that the asset in question adjusts less quickly to a change in net worth than to a change in interest rates. Unfortunately, while these statements can be made in principle, if we wish to retain the symmetry of the adjustment coefficient matrix, there is no way to estimate the separate elements of γ_1 and $\gamma_{\Delta W}$ other than to assume that one or the other of these vectors is the null vector.

I wish to conclude by correcting two errors in my original paper. The first was simply a matter of notation. For the sake of multiplying conformable matrices, equation (13Rb) of my paper should have appeared as

$$y_t = [\Gamma_{2,7},\ \Gamma_{8,12} - \gamma_{\Delta W} r',\ \gamma_1 + \gamma_{\Delta W}] \begin{bmatrix} X_t W_t \\ y_{t-1} \\ W_t \end{bmatrix} + V_t$$

The second error is more important and not so simple. In the second column on page 183 I wrote: "To assume that the elements of γ_1 are all equal to zero and all the elements of X_t are not constant over time implies that W_t does not enter equation system (1) as an isolated variable in one of the linear terms. This in turn implies there is no functional dependence of the elements of the vector of equilibrium values of assets and liabilities, y_i^*, on net worth, apart from a scale factor." In other words, to set the elements of γ_1 equal to zero is to set the elements of B_1 equal to zero in the relation:

$$\overset{*}{y_t} = (B_1 X_1 + B X_t) W_t$$

But this is not the case. The ith element of γ_1 is obtained by post-multiplying the ith row of the matrix A by the vector B_1 (whose elements, of course, sum to one). Thus, in addition to requiring that each column of A sum to zero, we require that each row of A be such that all such inner products sum to zero. We also note that even though the variable W does not appear in the equations, we can solve our estimate of A for an estimate of B_1 by rewriting equation (15a′). I wrote it in the paper as

$$\text{(15a}'\text{)} \qquad \begin{bmatrix} \hat{\Gamma}^a_{2,7} \\ 0 \end{bmatrix} = \begin{bmatrix} \hat{A}^a \\ r' \end{bmatrix} \hat{B}^a$$

where $\hat{\Gamma}^a_{2,7}$ was our estimate of the matrix AB, $\hat{A}^a$ our estimate of the matrix A, and the problem was to solve this system for $\hat{B}^a$ our estimate of B. We can rewrite this equation as:

$$\begin{bmatrix} 0 & \hat{\Gamma}^a_{2,7} \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} \hat{A}^a \\ r' \end{bmatrix} [\hat{B}^a_1 \hat{B}^a]$$

We have added a column to the matrix on the left and a column to the second matrix on the right[3] The additions merely state that each element of $\hat{A}^a\hat{B}^a_1$, our estimate of γ_1, equals zero, and that the elements of $\hat{B}^a_1$, our estimate of B_1, sum to unity. We can solve this system for $\hat{B}^a_1$ in the manner described in my paper for solving the system (15a′) for $\hat{B}^a$.

[3] It is instructive to compare this augmented version of equation (15a′) with my paper's equation (15b′) which referred to a system in which the elements of γ_1 were not constrained to zero, and the elements of $\hat{\gamma}_1$ are the regression coefficients of W.

$$\text{(15b}'\text{)} \qquad \begin{bmatrix} \hat{\gamma}_1 & \hat{\Gamma}^b_{2,7} \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} \hat{A}^b \\ r' \end{bmatrix} [\hat{B}^b_1 \hat{B}^b]$$

It is clear that the augmented version of (15a′) is the exact analogue of (15b′) for the case in which all the elements of γ_1 are constrained to zero.

REFERENCES

W. C. Brainard and J. Tobin, "Pitfalls in Financial Model Building," *Amer. Econ. Rev. Proc.*, May 1968, *58*, 99–122.

G. Chow, "On the Long-Run and Short-Run

The Industrial Composition of U.S. Export and Subsidiary Sales to the Canadian Market: Note

By H. G. Baumann*

In a recent article in this *Review*, Thomas Horst has shown that for *U.S.* firms, exports (X) and subsidiary sales (P) represent alternative means of utilizing their technological superiority over Canadian competitors. Furthermore, he showed that the Canadian tariff and relative cost factors determined whether the *U.S.* advantage was exploited through exports *or* subsidiary sales. The purpose of this note is to present some additional empirical support for Horst's conclusions, and to indicate areas where further research is desirable.

In the equations estimated below, two measures which refer to the output rather than the input of the technological effort of industries are employed. The substitution of research and development ($R\&D$) expenditures by patents issued in industry i in Canada (PA) yields the following results:[1]

$$X_i = .049 + \underset{(1.54)}{175.37}\, PA_i \qquad R^2 = .16 \quad \text{d.o.f.} = 12 \tag{1}$$

$$P_i = .172 + \underset{(3.09)}{594.15}\, PA_i \qquad R^2 = .44 \quad \text{d.o.f.} = 12 \tag{2}$$

$$S_i = .222 + \underset{(3.50)}{759.52}\, PA_i \qquad R^2 = .51 \quad \text{d.o.f.} = 12 \tag{3}$$

This compares with Horst's results for the same sample of industries as follows:

$$X_i = .032 + \underset{(2.55)}{6.05}\, R\&D_i \qquad R^2 = .35 \quad \text{d.o.f.} = 12 \tag{4}$$

$$P_i = .190 + \underset{(4.67)}{13.51}\, R\&D_i \qquad R^2 = .41 \quad \text{d.o.f.} = 12 \tag{5}$$

$$S_i = .221 + \underset{(4.11)}{19.56}\, R\&D_i \qquad R^2 = .58 \quad \text{d.o.f.} = 12 \tag{6}$$

Therefore, it appears questionable that technological intensity as measured by patents can explain *U.S.* exports to Canada (the critical value for the t-statistic at the 5 percent level with 12 degrees of freedom is 1.78), but that patents do better (higher R^2) than $R\&D$ in explaining subsidiary production. However, the former result changes dramatically when the following *log* formulation is used:

$$ln\ X_i = 5.17 + \underset{(3.82)}{.940}\, ln\ PA_i \qquad R^2 = .55 \quad \text{d.o.f.} = 12 \tag{7}$$

This formulation has a higher R^2 than the equivalent equation with the *log* of $R\&D$ as the explanatory variable, and in general the log form gives a slightly better fit. When *both* patents and $R\&D$ are introduced as independent variables the variables fail to be significant thus indicating a multicollinearity problem. It appears then that patents and $R\&D$ can be used interchangeably as measures of technological intensity.

However, some observers feel that patents, like $R\&D$, are undesirable as proxies for technological intensity. For example, there may be a bias from one industry to the next in terms of the extent to which patent protection is sought. Fortunately, McGraw-Hill has undertaken a survey over a number of years which asks managers to estimate the amount of sales of "new" products at some date in the future. New products (NP) are defined as products which are not being sold at the time of the survey. For example, in

* Assistant professor of economics, University of Western Ontario. I would like to acknowledge the helpful comments of my colleague Don McFetridge. Ed Kolano performed the computations.

[1] Precise definitions and data sources for the variables are given in the Statistical Appendix: $S_i \equiv X_i + P_i$.

1965, managers in the machinery industry estimated that 23 percent of 1969 sales would consist of products not yet produced. When these data are employed the estimated equations are:

$$X_i = -.021 + .913\, NP_i \quad (8)$$
$$(2.86)$$
$$R^2 = .45 \quad \text{d.o.f.} = 10$$

$$P_i = .356 + .154\, NP_i \quad (9)$$
$$(.181)$$
$$R^2 = .003 \quad \text{d.o.f.} = 10$$

$$S_i = .335 + 1.07\, NP_i \quad (10)$$
$$(1.14)$$
$$R^2 = .12 \quad \text{d.o.f.} = 10$$

The export equation confirms that the United States tends to exploit her comparative advantage in producing new products. The results are better than when $R\&D$ is the explanatory variable, which is not surprising given that NP adjusts for the efficiency of the $R\&D$ effort to some extent. A second reason for the higher R^2 with NP may be that the *U.S.* comparative advantage is probably not entirely based on a technological advantage, narrowly defined, but also on general organizational, marketing, and financial skills which are more crucial for new as compared with old products.[2]

The three equations above taken together give some support for the product cycle theory pioneered by Raymond Vernon. Apparently there is a lag between the production and export of a new product from the United States and the production of that same product by *U.S.* subsidiaries in Canada. This would account for the fact that production of *U.S.* subsidiaries does not appear to be concentrated in industries where new products form a large proportion of total sales. (By the time that a product is manufactured in Canada, i.e., after the imitation lag, it is already in the growth or mature phase of the product cycle, and hence, NP fails to be significant.) A more appropriate variable for the second equation would probably be the percentage of production carried on with new *processes* in the various industries, but these data are not available.[3] However, it is possible to adjust the $R\&D$ variable by weighting it with the percentage of entrepreneurs who believe that technological breakthroughs will occur in processing within a particular industry. The following results were obtained from this exercise:

$$P_i = .185 + 35.4\, RDA_i \quad (11)$$
$$(2.80)$$
$$R^2 = .44 \quad \text{d.o.f.} = 10$$

In conclusion, our results confirm those of Horst with the additional refinement that the comparative advantage of *U.S.* exports is largely based on product innovating while the comparative advantage of *U.S.* subsidiaries may be based on process innovating.

Horst also explains the choice between *U.S.* exports and foreign subsidiary production. In the Canadian case, he finds that the higher the Canadian tariff (t) and the smaller the size of the Canadian market (Z), the larger will be the ratio of *U.S.* exports to subsidiary production for any given industry. Although Horst mentions real cost considerations other than economies of scale as determining the above choice, he does not introduce them into his equations. However, existing data make it possible to introduce the ratio of *U.S.* to Canadian wages per man-hour paid (WA). Horst's original results, and results with this additional explanatory variable are given below:

$$ln\left(\frac{X_i}{S_i}\right) = -.23 - 2.45\, ln\,(1 + t_i) \quad (12)$$
$$(3.23)$$
$$- 12.6\, Z_i$$
$$(3.24)$$
$$R^2 = .70 \quad \text{d.o.f.} = .15$$

[2] This point was already made by Horst, but the following empirical results may substantiate his statement. The advertising to sales ratio in *U.S.* industries fails to be significant when introduced as an explanatory variable into the above equations. This variable may be taken as a proxy for the degree of product differentiation, and old and new products may be highly differentiated, but the former do not require innovative talents either of the scientific or managerial type.

[3] In a recent paper, David Ault has shown that the export performance of the steel industries of a sample of countries can be explained by their use of different steelmaking processes. Of course, new products and processes may occur simultaneously, and therefore this distinction is not always valid. One example of the latter case has been studied by J. E. Tilton.

$$(13)\quad ln\left(\frac{X_i}{S_i}\right) = \underset{(4.42)}{5.62 - 2.72\, ln\,(1 + t_i)} - \underset{(5.86)}{19.2\, Z_i} - \underset{(3.84)}{.046\, WA_i}$$
$$R^2 = .87 \quad \text{d.o.f.} = 14$$

Obviously Horst is on the right track since the R^2 is raised from .70 to .87 and the wage variable has the right sign and is significant. However, the model is still not complete since the costs of raw material and manufactured inputs have been left out. Conceivably one might construct a variable of input costs based on *U.S.* and Canadian prices for the major or representative inputs of the industries in the sample, but this would be a major undertaking.

Statistical Appendix

Horst was kind enough to supply me with data on the first six variables defined below. I would like to thank Imre Bernolak of the Department of Industry, Trade, and Commerce, Ottawa, for helping with the data on wages and material costs in *U.S.* and Canadian industries. Data are for the year 1963.

$P_i \equiv$ Sales by *U.S.*-owned subsidiaries in Canada divided by total sales of all Canadian firms plus total Canadian imports for commodity i.

$X_i \equiv$ *U.S.* exports to Canada divided by total Canadian sales plus total Canadian imports of commodity i.

$S_i \equiv X_i + P_i$

$R\&D_i \equiv$ Company sponsored research and development expenditures divided by total sales for industry i in the United States.

$t_i \equiv$ Estimated nominal protection of the Canadian industry.

$Z_i \equiv$ Ratio of Canadian production plus imports divided by *U.S.* domestic production for industry i.

$PA_i \equiv$ Patents granted divided by total shipments of industry i in Canada. Data for patents are from O. J. Firestone. Shipments data are from Statistics Canada. Patents are classified into industries according to their end use. For comparison purposes it may have been desirable to use *U.S.* data. However, only 5.5 percent of patents granted in Canada have a local origin while 69.8 percent originate in the United States and 24.7 percent come from elsewhere (based on total patents granted 1957, 1960, 1963). In any case, when only patents granted to *U.S.* firms or individuals are included, the empirical results do not change. Nor do they change if three-year totals (1957+1960+1963) are used which indicates that the flow in one year does in fact reflect the stock as Horst assumes.

$NP_i \equiv$ Percent of 1969 sales consisting of products which were not produced in 1965. The source for these data is McGraw-Hill Economics Department. Ideally, the data should refer to 1963, but experimentation with data from more recent years did not lead to results different from those presented in the note.

$RDA_i \equiv (R\&D)_i x A_i$ where $A \equiv$ fraction of managers who believe that important breakthroughs will occur in processing as opposed to new products in industry i. The sources for the data are the same as those given for RD_i and NP_i above, respectively.

$WA_i \equiv$ *U.S.* wages per man-hour divided by Canadian wages per man-hour. Data for this variable are from Department of Industry, Trade, and Commerce, Ottawa. The data were converted from a three-digit basis to a two-digit basis by using employment figures as weights.

REFERENCES

D. Ault, "The Determinants of World Steel Exports: An Empirical Study," *Rev. Econ. Statist.*, Feb. 1972, *54*, 38–46.

O. J. Firestone, *Economic Implications of Patents*, Ottawa 1971.

T. Horst, "The Industrial Composition of U.S.

Exports and Subsidiary Sales to the Canadian Market," *Amer. Econ. Rev.*, Mar. 1972, *62*, 37–45.

J. E. Tilton, *International Diffusion of Technology: The Case of Semiconductors*, Washington 1971.

R. Vernon, "International Investment and International Trade in the Product Cycle, *Quart. J. Econ.*, May 1966, *80*, 190–207.

Department of Industry, Trade, and Commerce, *Comparative Tables of Principal Statistics and Ratios for Selected Manufacturing Industries: Canada and United States*, Ottawa 1971.

McGraw-Hill Economics Department, *Research and Development in American Industry*, New York 1966.

Statistics Canada, *Manufacturing Industries of Canada: Summary for Canada, 1963*, Ottawa 1966.

Lifetime Portfolio Selection in Continuous Time for a Multiplicative Class of Utility Functions

By GORDON PYE*

In a recent paper, Richard Meyer has studied the lifetime portfolio problem in continuous time for a multiplicative class of utility functions. Though obtaining a number of general characteristics of the optimal policy, he is able to obtain an analytic solution only for a special limiting case that corresponds to the additive family. The solution for this additive family is known from the independent work of Robert Merton, who, for continuous time, attacked this case directly. The continuous time case requires the assumption that the distribution of returns follow an infinitely divisible normal process. The discrete time case does not require this assumption. The additive family in discrete time has been studied by Edmund Phelps, Nils Hakansson, David Levhari and T. N. Srinivasan, and Paul Samuelson.

The purpose of this note is to point out that an analytic solution can be obtained for a subclass of the multiplicative family studied by Meyer. For discrete but not continuous time, I have given this solution in my 1972 paper. This solution differs significantly from that for the additive case in either discrete or continuous time. In the additive case, for a stationary distribution of returns, the proportion of wealth invested in risky securities doesn't change with age. This is not true for the solution in the multiplicative case. The proportion of wealth invested in risky securities increases or decreases with age as risk aversion is greater or less than that of the logarithmic utility function. The measure of risk aversion is the index of relative risk aversion developed by Kenneth Arrow and by John Pratt. If the logarithm may be thought of as the dividing line between optimists and pessimists, this result may be interpreted as follows. Optimists tend to gamble less as they grow older as they have less to gain, whereas pessimists gamble more as they have less to lose. Unlike the additive case, risk aversion in portfolio selection also depends on impatience in the multiplicative solution. Risk aversion in portfolio selection increases with impatience for optimists, while the reverse holds for pessimists. As the propensity to consume also depends on impatience, this means risk aversion and the propensity to consume will be correlated through their mutual dependence on impatience. For optimists, the correlation will be positive, whereas for the pessimists, it will be negative.

Consider an investor at time t with a remaining lifespan of $T-t$. Following Meyer, the utility of his remaining consumption stream will be taken to be the following:

$$(1)\quad U_t = \delta \exp\left\{\gamma \int_t^T \alpha^s u(c(s))ds\right\}$$
$$= \lim_{\Delta t\to 0} \delta \exp\left\{\gamma\alpha^t u(c(t))\Delta t\right\} U_{t+\Delta t}$$
$$\delta = \begin{cases} +1 & \gamma > 0 \\ -1 & \gamma < 0 \end{cases}$$

In (1), $u(c)$ is the utility of consuming at the rate c at any instant of time, α is a discount factor reflecting impatience which satisfies $0<\alpha<1$, and γ is a parameter which will turn out to be related to risk aversion in portfolio selection in the solution to be obtained below. It is to be noted that the additive family studied by Merton can be obtained from (1) by expanding in terms of γ, letting γ approach zero, and disregarding terms of order γ^2 or smaller.

The investor's net worth at t is denoted by w_t. There are two securities available for investment, one risky and the other riskless. Let z_{1t} be the proportion of net worth invested in the risky security, and let $1-z_{1t}$ be

* Professor of business administration, University of California, Berkeley.

the proportion invested in the riskless one. The risky return is taken to be infinitely divisible and normally distributed. Denoting the return over Δt by $r_t(\Delta t)$, the mean and variance of $r_t(\Delta t)$ are taken to be $\mu\Delta t$ and $\sigma^2\Delta t$ for any Δt. The return on the riskless security is taken to be $\rho\Delta t$. It is assumed that $\mu > \rho$. The proportionate rate at which net worth is consumed will be denoted z_{0t}. The consumption rate is therefore: $c(t) = z_{0t}w_t$.

Let $V(w_t, t)$ denote the maximum expected value of U_t, given w_t, and that an optimal policy is followed in t and subsequently. Use of (1) and application of the optimality principle gives the following relation which must be satisfied by V for small Δt:

$$(2) \quad V(w_t, t) = \underset{z_{0t}, z_{1t}}{\text{Max}} \exp\left\{\gamma\alpha^t u(z_{0t}w_t)\Delta t\right\} E\left\{V(w_{t+\Delta t}, t + \Delta t) \mid w_t\right\}$$

For small Δt, the dependence of investment income on consumption withdrawal over Δt can be neglected so that $w_{t+\Delta t}$ is given as follows:

$$(3) \quad w_{t+\Delta t} = (1 + \rho\Delta t + (r_t(\Delta t) - \rho\Delta t)z_{1t} - z_{0t}\Delta t)w_t$$

Expanding the exponential in (2) with respect to Δt; expanding $V(w_{t+\Delta t}, t+\Delta t)$ with respect to both its arguments about w_t and t, respectively; and making use of (3) and the properties of $E\{r_t(\Delta t)\}$ and $\text{Var}\{r_t(\Delta t)\}$; one obtains for small Δt (neglecting terms of order $(\Delta t)^2$ or less):

$$\begin{aligned}(4) \quad V(w_t, t) = \underset{z_{0t}, z_{1t}}{\text{Max}}\, & V(w_t, t) \\ & + (\rho + (\mu - \rho)z_{1t} \\ & - z_{0t})\Delta t w_t V_w(w_t, t) \\ & + \Delta t V_t(w_t, t) \\ & + 1/2\, \sigma^2 \Delta t z_{1t}^2 w_t^2 V_{ww}(w_t, t) \\ & + \gamma\alpha^t u(z_{t0}w_t)\Delta t V(w_t, t)\end{aligned}$$

where V_w, V_{ww}, and V_t have been used to denote partial derivatives with respect to the first and second arguments of V. Dropping $V(w_t, t)$ from both sides of (4), dividing by Δt, suppressing the t subscript on the state and decision variables, and suppressing the arguments of V and its partial derivatives gives:

$$\begin{aligned}(5) \quad 0 = \underset{z_0, z_1}{\text{Max}}\, & (\rho + (\mu - \rho)z_1 - z_0)wV_w + V_t \\ & + 1/2\, \sigma^2 z_1^2 w^2 V_{ww} + \gamma\alpha^t u(z_0 w)V\end{aligned}$$

The optimal policy must satisfy the partial differential equation given in (5). The equation in (5) agrees with that given by Meyer with the following exceptions: V_t is not taken to be equal to zero even when $T = \infty$, since V and the optimal policy depend on t; α is not suppressed by setting $\alpha = 1$ since having $0 < \alpha < 1$ is necessary for the proper covergence of V when $T = \infty$; and the sign of γ is not assumed to be negative, as having risk aversion either increasing or decreasing with age is taken to be a logical possibility.[1] The equation in (5) agrees with that given by Merton except that the last term containing $u(c)$ is multiplied by V, reflecting the fact that the multiplicative and not the additive family is being considered.

When $u(c) = \ln c$, an analytic solution can be obtained to (5) for either a finite or an infinite horizon. The latter can be obtained directly or as a limiting form of the finite case, providing $0 < \alpha < 1$ as is being assumed. For simplicity, only the solution to the infinite case will be derived here. Portfolio selection behavior in either case is of the same form. The solution is given in the following theorem:

THEOREM: *The solution to the lifetime portfolio selection problem in continuous time for* $u(c) = \ln c$ and $T = \infty$ *is as follows:*

$$\begin{aligned}(6) \quad V &= a(t)w^{\lambda(t)} \qquad \lambda(t) = -\gamma\alpha^t/\ln\alpha \\ a(t) &= \delta(-\ln\alpha)^{\lambda}\exp\left\{-\lambda(1 + \rho/\ln\alpha)\right\} \cdot (1 - \lambda)^{(\mu-\rho)^2/2\sigma^2 \ln\alpha}\end{aligned}$$

[1] It is to be noted that having $\alpha = 1$, $T = \infty$, and $\gamma < 0$, as does Meyer, will mean necessarily that $V = 0$. This follows because a policy of investing all net worth in the riskless asset and consuming the riskless income will make $\int_t^\infty u(c(s))ds$ unbounded with certainty and hence:

$$V = E\left\{\exp\left\{-\int_t^\infty u(c(s))ds\right\}\right\} = 0$$

Presumably, many other policies will also give $V = 0$, so the problem is not well formulated.

$$\overset{*}{z}_{0t} = \overset{*}{z}_0 = -\ln\alpha$$

$$\overset{*}{z}_{1t} = (\mu - \rho)/\sigma^2(1 - \lambda)$$

To see that (6) is the solution to (5) under the specified conditions, substitute $u(c) = \ln c$ and the V given in (6) into (5). Using $wV_w = \lambda V$, $w^2 V_{ww} = \lambda(\lambda - 1)V$, and $V_t = \dot{\lambda} V \ln w + \dot{a}V/a$ (where the dot notation indicates the time derivative) gives:

$$(7) \quad 0 = \underset{z_0, z_1}{\operatorname{Max}} V[(\rho + (\mu - \rho)z_1 - z_0)\lambda + \dot{\lambda}\ln w + \dot{a}/a + 1/2\,\lambda(\lambda - 1)\sigma^2 z_1^2 + \gamma\alpha^t \ln(z_0 w)]$$

Taking δV outside the maximization operator because it is positive and independent of z_0 and z_1, and observing that $\dot{\lambda}\ \ln w + \gamma\alpha^t \ln w = 0$, gives the following:

$$(8) \quad 0 = \underset{z_0, z_1}{\operatorname{Max}} \delta[\lambda(\rho + (\mu - \rho)z_1 - z_0) + \dot{a}/a + 1/2\,\lambda(\lambda - 1)\sigma^2 z_1^2 + \gamma\alpha^t \ln z_0]$$

$$0 = \dot{a}/a + \lambda\rho + \lambda \underset{z_0}{\operatorname{Max}} \{-z_0 - \ln\alpha \ln z_0\} + \lambda \underset{z_1}{\operatorname{Max}} \{(\mu - \rho)z_1 - 1/2\,(1 - \lambda)\sigma^2 z_1^2\}$$

Solution of these two simple maximization problems gives $z_0^* = -\ln\ \alpha$ and $z_{1t}^* = (\mu - \rho)/\sigma^2(1 - \lambda)$, as stated in (6). It is to be noted that the maximum with respect to z_1 has been taken as internal and independent of constraints on margin purchases or short sales. Substitution of these solutions into (8) shows that V is the required solution, if $\dot{a}/a$ satisfies the following differential equation:

$$(9) \quad 0 = \dot{a}/a + \lambda\rho + \gamma\alpha^t(\ln(-\ln\alpha) - 1) + \lambda(\mu - \rho)^2/2\sigma^2(1 - \lambda)$$

By using the $a(t)$ given in (6), it is easily verified that this $a(t)$ does provide a solution to (9).

For finite T, the portfolio solution will be of the same form as (6) except that $\lambda = -\gamma\alpha^t(1 - \alpha^{T-t})/\ln\ \alpha$. The optimal propensity to consume is no longer independent of t, being given by:

$$\overset{*}{z}_{0t} = -\ln\alpha/(1 - \alpha^{T-t})$$

The expression for $a(t)$ is considerably more cumbersome.

As shown in the introduction, the limiting case as $\gamma \to 0$ for (1) gives the additive family. Here, the special case of the additive family with a logarithmic utility function is obtained. The solution for this has been given by Merton. It corresponds to the solution in (6) for z_0^* and z_{1t}^* with $\gamma = 0$ (i.e., $\lambda = 0$).

The solution in (6) for z_{1t}^* is easily interpreted to verify the results stated in the introduction. Differentiating z_{1t}^* with respect to t gives:

$$(10) \quad \frac{d\overset{*}{z}_{1t}}{dt} = -\gamma\alpha^t(\mu - \rho)/\sigma^2(1 - \lambda)^2$$

Thus, the proportion of the portfolio invested in the risky security increases with age if $\gamma < 0$ and decreases with age if $\gamma > 0$. The additive case with $\gamma = 0$ forms the watershed. Referring back to (5), this is easily related to the Arrow-Pratt index of relative risk aversion in portfolio selection. The optimal portfolio at any time is that obtained by maximizing $E\{V\}$. This may be seen as follows. For gambles, small relative to wealth, $E\{V\}$ is equal to a linear function of mean and variance where the coefficient of variance is $wV_{ww}/2V_w$, or $-1/2$ times the index of relative risk aversion. Here the gamble will be small, relative to wealth for small Δt. From (5) the optimal portfolio is given by

$$wV_w \underset{z_1}{\operatorname{Max}} \left\{\rho + (\mu - \rho)z_1 + \frac{wV_{ww}}{2V_w}\sigma^2 z_1^2\right\}$$

where the term in brackets is the required linear function of mean and variance of return. From the V given in (6) the index of relative risk aversion in portfolio selection at any time is $1 - \lambda(t)$. It will be greater or less than that of the logarithm (which is one) as γ is negative or positive. Thus, risk aversion in portfolio selection as measured by

this index is seen to rise or fall with age as risk aversion is less or greater than that of the logarithm.

Differentiating z^*_{1t} with respect to α gives

$$(11) \quad \frac{dz^*_{1t}}{d\alpha} = \gamma\alpha^{t-1}(\mu - \rho)(1 - t\, ln\, \alpha) / \sigma^2(1-\lambda)^2(ln\, \alpha)^2$$

Thus the proportion invested in the risky security increases or decreases with increasing patience as risk aversion is less or greater than that of the logarithm. The propensity to consume (i.e., $z^*_0 = -ln\, \alpha$) decreases with increasing patience. Therefore, risk aversion in portfolio selection and the propensity to consume will be negatively or positively correlated, as risk aversion is less or greater than that of the logarithm.

REFERENCES

K. J. Arrow, *Aspects of the Theory of Risk Bearing*, Helsinki 1965.

N. H. Hakansson, "Optimal Investment and Consumption Strategies Under Risk for a Class of Utility Functions," *Econometrica*, Sept. 1970, *38*, 587–607.

D. Levhari and T. N. Srinivasan, "Optimal Savings Under Uncertainty," *Rev. Econ. Stud.*, Apr. 1969, *51*, 153–64.

R. C. Merton, "Lifetime Portfolio Selection Under Uncertainty: The Continuous Time Case," *Rev. Econ. Statist.*, Aug. 1969, *51*, 247–57.

R. F. Meyer, "On the Relationship Among the Utility of Assets, the Utility of Consumption, and Investment Strategy in an Uncertain but Time-Invariant World," in J. Lawrence, ed., *Proc. Fifth International Conference on Operational Research*, Tavistock 1970.

E. S. Phelps, "The Accumulation of Risky Capital: A Sequential Utility Analysis," *Econometrica*, Oct. 1962, *30*, 729–43.

J. Pratt, "Risk Aversion in the Small and in the Large," *Econometrica*, Jan. 1964, *32*, 122–36.

G. Pye, "Lifetime Portfolio Selection with Age Dependent Risk Aversion," in G. P. Szegö and K. Shell, eds., *Mathematical Methods in Investment and Finance*, North Holland 1972.

P. A. Samuelson, "Lifetime Portfolio Selection by Dynamic Stochastic Programming," *Rev. Econ. Statist.*, Aug. 1969, *51*, 239–46.

Social Return to Public Information Services: Statistical Reporting of U.S. Farm Commodities: Comment

By LAWRENCE N. SMITH AND BRUCE A. SCHERR*

In a recent issue of this *Review*, Yujiro Hayami and Willis Peterson present a practical application of Alfred Marshall's social welfare concepts.[1] Their approach, in the tradition of public goods analysis, seeks to measure the social returns to improvements in information about *U.S.* farm commodities. Their strong theoretical argument is jeopardized by their empirical evidence leading to exaggerated and unrealistic conclusions. We have two basic criticisms of their analysis: 1) they use highly inelastic demand elasticities for grains which exclude export and animal feed uses; and 2) their conclusions are dominated by the marginal social returns for only two of the seventeen commodities studied. Their distribution of returns raises specific questions about sampling techniques for the two commodities which may be more important than the broader issue of social returns to public information services.

I. Erroneous Use of Some Elasticities

The price elasticities of demand used by Hayami and Peterson for the three food grains and three feed grains create exaggerated returns and highly sensitive benefit-cost ratios of meaningless dimensions for practical government decision making. Government decision making, using benefit-cost ratios, requires both a "yes we do" (or "no we don't") decision as well as a "how far do we go" decision. The Hayami and Peterson analysis appears to suggest that government should invest in sampling improvements to reduce typical sampling error for all commodities to the 0.0 percent level. For a 0.0 percent sampling error complete enumeration is required which we are sure was not the intent of the analysis.

The logic of the theoretical argument presented is that the more inelastic the demand function for a given commodity, the greater is the potential for increased social returns. The elasticities, taken from G. Brandow, p. 59, range from −.01 to −.04 for oats, barley, wheat, corn, rye, and rice. Brandow specifically states that these elasticities exclude the quantities demanded for export or for feed.

Using Hayami and Peterson's method, the 15.8 percent sampling error in reporting annual rice production could set off an inventory holding action which would last until the price of rice increased 395 percent to $24.75 per cwt. The inflated price according to their estimation, would cause a social loss of $289.6 million as the error is allowed to run its course. In all important applications of the concept of consumer surplus, Marshall concerned himself solely with the range of normal price variation.[2] Furthermore, historically (last 15 years) the greatest annual change in rice prices was $.59 (13 percent) and the average annual change in price was only 2.6 percent.

For the years included in the Hayami and Peterson study (1966–68) 9 percent of *U.S.* corn production was used for human food and industrial use and only 5 percent of the oats. Thirty percent of the barley produced was used for human food and industrial use.[3] This indicates that the demand elasticities used should be for combined uses, and that social loss, as defined by Hayami and Peterson, should be figured on a crop value for human food and industrial use for the six food grains.

Brandow, in the same reference used by Hayami and Peterson states, "... rough estimates place the elasticity for low-protein

* President, Lawnel Farms, Inc., Piffard, N.Y. and graduate instructor, Purdue University.

[1] See Marshall, pp. 124–133, 140, and 810–812.

[2] See Mark Blaug, p. 262.

[3] See *Agricultural Statistics 1969*.

TABLE 1—A COMPARISON OF ESTIMATES OF MARGINAL SOCIAL BENEFIT-COST RATIOS CORRESPONDING TO REDUCTION IN TYPICAL SAMPLING ERROR IN THE SURVEY FOR STATISTICAL REPORTING OF FARM COMMODITIES, UNITED STATES, FROM HAYAMI AND PETERSON TO SMITH AND SCHERR, 1966–68.

Change in Typical Sampling Error (percent)	Hayami-Peterson		Smith-Scherr	
	Area Sample	Multiple Frame Sampling	Area Sample	Multiple Frame Sampling
from 3.0–2.5	824.0	824.0	312.0	312.0
2.5–2.0	619.0	619.0	238.0	238.0
2.0–1.5	93.0	106.0	39.1	44.4
1.5–1.0	57.0	60.0	22.5	23.7
1.0–0.5	7.4	13.0	2.8	4.8
0.5–0.0	1.0	1.6	0.3	0.6

feeds at −.36." He further suggests that the demand for corn for all purposes apparently has an elasticity of about −.50, and that the elasticities for the other feed grains could considerably exceed unity because they comprise so small a share of the total low protein feed supply.[4] Earlier work by Henry Schultz suggested an elasticity coefficient for corn at the farm level of −.77, for oats −.54 and for barley −.17. Recent work at Purdue by Peter Emerson (p. 155) found a price elasticity of demand for wheat for all uses of −.25 and for rice −.20.

It is not the purpose of this comment to engage in an exchange of contrary estimates of price elasticity of demand. With adjustments in demand elasticities for combined use, we found a 67 percent reduction in the total benefits generated when the sampling error is improved from 3.0 percent to 2.5 percent. Table 1 shows the more reasonable benefit-cost ratios which emerge when the elasticities referenced above for combined use for the six grains are used.

The new lower levels of return still fall within the 500 percent range allowed for inaccurate coefficients by Hayami and Peterson. Can error in excess of 1000 percent for the six grains be considered "good enough for government work"? The lower levels of return representing new lower bounds than those reported by Hayami and Peterson still appear to support arguments for improved crop information. The higher elasticities have the added advantage of lending credence to the benefit-cost ratios indicating a cutoff point short of complete enumeration at the 0.5 percent typical sampling error level.

II. Dominance of Rice and Potatoes

The Hayami and Peterson justification for improving public information services is largely based on rice and potatoes. Table 2 shows results using our adjusted elasticities when rice and potatoes are removed from the analysis. Before adjustment, 56 percent of the marginal social benefits from improved information came from rice and potatoes. If the elasticity for rice is adjusted upward as we suggest, then potatoes alone provide 54 percent of the marginal social returns. At the time of the 1964 agricultural census, rice made up .6 percent and potatoes .4 percent of U.S. total crop acres. Perhaps the research question should have been, "Should the government invest to improve crop information about potatoes?"

In the public goods tradition we ask, "Can the social returns to potato eaters be used to justify, as their analysis shows, a complete enumeration of all commodities?" With such a strikingly skewed distribution of returns, it would appear more appropriate to ask potato eaters and the potato industry to pay for the cost of their information knowing that benefits to the consumers of the other commodities are small, making it possible for benefits to be divided.

A more practical approach for Hayami

TABLE 2—COMPARISON OF MARGINAL BENEFIT-COST RATIOS AS CALCULATED BY SMITH AND SCHERR, CORRESPONDING TO CHANGES IN TYPICAL SAMPLING ERRORS, UNITED STATES 1966–68

Change in Typical Sampling Error (percent)	Marginal Benefit-Cost Ratio			
	With Rice and Potatoes		Without Rice and Potatoes	
	Area Sample	Multiple Frame Sampling	Area Sample	Multiple Frame Sampling
from 3.0 to 2.5	312.0	312.0	87.2	87.2
2.5 to 2.0	238.0	238.0	68.1	68.1
2.0 to 1.5	39.1	44.4	9.5	10.8
1.5 to 1.0	22.5	23.7	5.8	6.1
1.0 to 0.5	2.8	4.8	0.7	1.2
0.5 to 0.0	0.3	0.6	0.0	0.1

[4] This argument is supported by Brandow, p. 8.

TABLE 3—PROPORTIONED SURVEY COST AND MARGINAL BENEFIT-COST RATIOS FOR SIX GRAINS, 1966–68

	Six Grains[a] (million dollars)						
Area Sample	1.2	1.32	1.45	2.04	2.78	6.03	21.88
Multiple Frame Sample	1.2	1.32	1.45	1.97	2.68	4.58	15.60
Typical Sampling Error, Major Commodities	3.0	2.5	2.0	1.5	1.0	0.5	0.0

Change in Typical Sampling Error (percent)	Marginal Survey Cost: Area	Marginal Survey Cost: Multiple Frame	Grains Inventory Adjustment (million dollars)	Grain Returns Relative to Grain Costs, Marginal Benefit-Cost Ratios: Area	Grain Returns Relative to Grain Costs, Marginal Benefit-Cost Ratios: Multiple Frame
from 3.0 to 2.5	0.12	0.12	27.1	225.0	225.0
2.5 to 2.0	0.13	0.13	19.1	146.0	146.0
2.0 to 1.5	0.59	0.52	11.9	20.0	22.0
1.5 to 1.0	0.74	0.71	9.5	12.8	13.3
1.0 to 0.5	3.35	1.90	5.5	1.6	2.8
0.5 to 0.0	15.85	11.02	3.3	0.2	0.3

[a] $\frac{6\text{ (Survey Cost)}}{17} = \text{Proportioned Cost}$

and Peterson would have been to place potatoes in their production adjustment model with more reasonable marginal social returns as a result. This approach could be justified since potatoes can be grown somewhere in the United States each season of the year.

The Hayami-Peterson analysis does not show the relationship between survey costs for each commodity and returns for each commodity. It should be apparent that crops covering less than 1 percent of total crop area would cost less to survey than the major commodities. We have prepared a table (Table 3) which uses survey costs proportioned evenly among commodities to arrive at benefit-cost comparisons for the six grains. Our results show a definite investment cutoff point at the 0.5 percent typical sampling error level using a stratified or multiple frame sampling method.

III. Conclusions

In conclusion we support the theoretical argument used by Hayami and Peterson if it is limited to the normal range of commodity price changes. We support efforts to improve farm commodity production information on the basis of adjusted elasticities and more specific benefit-cost relationships. Our major regret is that the article presented a careless empirical analysis, leading to meaningless benefit-cost ratios which are useless, if not misleading, for public decision making. Presumably, to provide guidelines for such decision making is what Hayami and Peterson's efforts—if not efforts of all economists—are supposed to accomplish.

REFERENCES

M. Blaug, *Economic Theory in Retrospect*, 2d ed., Homewood 1968.

G. E. Brandow, *Interrelations Among Demand for Farm Products and Implications for Control of Market Supply*, Agr. Exp. Sta. Bull. 680, Penn. State Univ. 1961.

P. M. Emerson, "An Economic Analysis of World Corn Trade," unpublished doctoral dissertation, Purdue Univ. 1972.

Y. Hayami and W. Peterson, "Social Returns to Public Information Services: Statistical Reporting of U.S. Farm Commodities," *Amer. Econ. Rev.*, Mar. 1972, *62*, 119–30.

A. Marshall, *Principles of Economics*, 8th ed., London 1916.

H. Schultz, *The Theory and Measurement of Demand*, Chicago 1938.

U.S. Department of Agriculture, *Agricultural Statistics 1969*, Washington 1969.

Social Return to Public Information Services: Statistical Reporting of U.S. Farm Commodities: Reply

By YUJIRO HAYAMI AND WILLIS PETERSON*

We agree with Lawrence Smith and Bruce Scherr that the demand elasticities should apply to the total market demand for the commodities in question rather than to one particular market. However the choice of demand elasticities is not quite as simple as they might lead us to believe. It is necessary to keep in mind that the time allowed for quantity to adjust to price changes in the inventory adjustment model is relatively short. The adjustment period implicit in this model takes place within the crop year. In the case of feed grains, the quantity adjustment has to come largely from changes in the rate of feeding rather than from changes in livestock numbers. Since changes in the feeding rate generally are profitable over only a very narrow range for most classes of livestock, we would expect the short-run demand elasticities to be substantially less than long-run elasticities which allow for adjustments in livestock and poultry numbers.

The demand elasticities presented by G. Brandow which our critics would have us use are long-run elasticities that were applied to a long-run supply control model. As Brandow specifically states, "Adjustments thus described for the feed-livestock economy are long term and would take place with varying lags, probably coming most slowly in dairying" (p. 7). In regard to the elasticities, he points out, "The elasticities used in the model are intended to apply when enough time is allowed for livestock adjustments to take place and for *inventory changes to be unimportant*" (p. 9, italics added for emphasis). Although the authors do not specify whether they use Brandow's estimates or not, it is fairly clear that they use comparable figures in their calculations. Thus they apply long-run demand elasticites to a short-run adjustment model, which to us does not seem entirely appropriate either.

Although we welcome efforts to estimate demand and supply elasticities specifically applicable to the models at hand, we do not feel that much can be gained by searching the literature for parameters that may bias the estimates one way or another. The results are not likely to change the overall conclusions of the study. Indeed the estimates presented by Smith and Scherr fall within the range of estimates presented in Table 5 of our article. This table is not intended to bracket the true values, as our critics imply, but mainly to illustrate that the marginal benefit-cost (B/C) ratios of increasing the overall average sampling accuracy from 2.0 to 1.5 percent remains relatively attractive even when we impose these rather severe downward biases on the data. The Smith-Scherr estimates show the same thing. Certainly we do not suggest that the sampling error be reduced to zero. When dealing with such gross figures as this, a B/C ratio in the neighborhood of one certainly would not be considered attractive. However, it should be stressed that the estimates presented in Table 4 of the original article already contain a downward bias in that the inventory adjustment model is applied only to crops and the production adjustment model is applied only to livestock.[1]

In reference to the dominance of rice and potatoes in the social returns, we should point out that the article deals only with a specific and fairly narrow range of sampling errors. Within a different range, such as between 5 and 10 percent, we might reasonably expect other commodities such as wheat and corn to provide the major share of the

* Professors, Tokyo Metropolitan University, and University of Minnesota, respectively.

[1] We should also point out that an additional social return is likely to be forthcoming if the government can use such information to obtain a more efficient allocation of its investment expenditures.

returns. Of course, if data were available on the relationshlp between sampling cost and error of specific commodities the B/C ratios could be estimated for each commodity separately. We did not do this because the data were not available to us.

In regard to the wide range of price changes implied by the sampling errors and demand elasticities, it is necessary to bear in mind that the sampling errors are averages of errors inherent in a number of surveys conducted from planting to harvest. Naturally the expected price would come closer to the market clearing price as harvest time approaches. As an analogy, it is considerably easier to predict the score of a football game in the fourth quarter than before the game begins. Given sufficient information, it would be possible to calculate the social returns of additional information that is made available during the crop year.

We would hope that this comment and reply does not detract from what is perhaps the most important aspect of the article, namely that investment in market information can be viewed as yielding a social return to society in general as opposed to only a means of increasing short run profits to primary producers or marketing firms.

REFERENCES

G. E. Brandow, *Interrelations Among Demand for Farm Products and Implications for Control of Market Supply,* Agr. Exp. Sta. Bull. 680, Penn. State Univ. 1961.

On Ignorance and Equal Distribution

By AMARTYA SEN*

There has recently been a revival of interest in Abba Lerner's proposition on maximization of probable aggregate welfare through a policy of equal distribution in a situation of Bayesian ignorance as to who has which utility function. Maurice McManus, Gary Walton, and Richard Coffman, and Roger McCain have presented in the June 1972 issue of this *Review* formal versions of Lerner's theorem based on explicit formulation of probability distributions of utility functions for different individuals. Further, McManus et al. have shown that the optimality of equal distribution holds also under the "maximin" strategy of maximizing the lowest value of aggregate welfare for any distribution, and this result does not, naturally, depend on the exact assumptions about probability distributions. As it happens, these two theorems were stated and proved in my paper presented to the International Economic Association Round-Table Conference on Public Economics in Biarritz in 1966 and were published in Sen (1969) as Theorems II ("Probabilistic Egalitarianism") and III ("Maximin Egalitarianism").

This fact in itself is scarcely interesting, except conceivably to the authors and editors, but what may interest the readers of this journal is the fact that both these theorems can be substantially generalized by dropping the framework of *additive* social welfare functions and only requiring social welfare to be a concave (or quasi-concave) function of individual utilities, which need not even be separable.

This generalization is of particular relevance in the context of Milton Friedman's penetrating criticism of Lerner's approach:

> Eliminate the assumption of ignorance, and the same analysis immediately yields a justification of inequality if individuals do differ in capacity to enjoy satisfaction. . . . Suppose, further, that it is discovered by this technique that a hundred persons in the United States are enormously more efficient pleasure machines than any others, so that each of these would have to be given an income ten thousand times as large as the income of the next most efficient pleasure machine in order to maximize aggregate utility. Would Lerner be willing to accept the resulting division of income as optimum . . . ? [pp. 310–11]

The theorems presented below indicate that Lerner does not have to do any such thing since the maximizing property of the equal distribution holds for probable social welfare even if it is not of the additive form. A concave social welfare function W need not necessarily award a higher income to a more efficient pleasure machine.[1] The "maximin" result presented in Sen (1969) and McManus et al. can also be similarly generalized.

I

The results are proved for the case in which there are n individuals and n utility functions, when it is not known who has which utility function.[2] The following notation is used:

y = any income vector $(y_1, \ldots, y_n)$, in which y_i is the income of person i

z = vector of equal incomes, i.e., $z_i = z_j$ for all i, j

U^j = individual welfare function j

W = social welfare

u_i = individual welfare of person i

p_i^j = probability of person i having U^j

* Professor of economics, London School of Economics.

[1] This is an important issue for the measurement of income distribution, on which see Partha Dasgupta, Sen, and David Starrett, and Michael Rothschild and Joseph Stiglitz. I have discussed the underlying normative framework in my forthcoming book. The "Weak Equity Axiom" used there would guarantee that the more efficient pleasure machine would get, in fact, *less* income, and this is consistent with a concave social welfare function. See also Sen (1970), ch. 9.

[2] We can take a wider set V of individual welfare functions and extend the proofs. On this see Sen (1969).

ASSUMPTION 1 (*Total Income Fixity*): The total income to be distributed is constant (y^*):

$$(1) \qquad y_1 + \ldots + y_n = y^*$$

ASSUMPTION 2 (*Concavity of the Group Welfare Function*): Social Welfare W, an increasing and symmetric function of individual welfare levels: $W = W(u_1, \ldots, u_n)$, is concave.

ASSUMPTION 3 (*Concavity of Individual Welfare Functions*): Each individual welfare function $U_j(y)$ is concave.

ASSUMPTION 4 (*Equi-probability*): For each j, $p_i^j = p_m^i$, for every pair of individuals i, m.

THEOREM 1: *Given Assumptions 1, 2, 3 and 4, the mathematical expectation of social welfare is maximized by an equal division of income.*

PROOF:

Thanks to the symmetry property of W, we can take the social welfare function to be $W(u^1, \ldots, u^n)$, where u^j is the utility of the person with the jth utility function; nothing depends on precisely which numbered individual enjoys U^j. Corresponding to any income distribution vector y, let $\bar{y}$ be a permutation of it such that $\bar{y}^j$ is the income going to the person with the jth utility function. There are $n!$ different ways of assigning n individual welfare functions to n persons, and corresponding to each such assignment k, we have a permuted vector $\bar{y}(k)$, for any income vector y. Therefore for any distribution y, there are $n!$ values of social welfare given by $F(\bar{y}(k))$, for $k = 1, \ldots, n!$, and given Assumption 4, the expected value of social welfare from y is given by:

$$(2) \qquad E(y) = \frac{1}{n!} \sum_k F(\bar{y}(k))$$

Of course, since all permutations z of the equi-distribution vector z are the same:

$$(3) \qquad E(z) = F(z)$$

By Assumptions 2 and 3, F is a concave function, and by Assumption 1, we have:

$$(4) \qquad z = \frac{1}{n!} \sum_k \bar{y}(k)$$

Since F is concave, it is clear that:

$$(5) \qquad E(y) \leq E(z)$$

Since this is true of all y, Theorem 1 holds.

II

For the "maximin" result, Assumption 4 is weakened to:

ASSUMPTION 4* (*Shared Set of Welfare Functions*): For any individual i and any welfare function j, it is possible that i has j. Nothing is said on relative probabilities.

For the existence of minimal individual welfare levels, something has to be assumed about the lower bounds of these functions. The following simple assumption is sufficient.[3]

ASSUMPTION 5 (*Bounded Individual Welfare Functions*): Each individual welfare function U^j is bounded from below.

THEOREM 2: *Given Assumptions 1, 2, 3, 4*, and 5, the maximin strategy for social welfare is to distribute income equally.*

PROOF:

Since z is an average of $\bar{y}(k)$ for all k, for any quasi-concave welfare function F, we must have:

$$(6) \qquad F(z) \geq \operatorname*{Min}_k F(\bar{y}(k))$$

Furthermore, z being an equal division, the value of $\bar{z}$ is invariant with respect to interpersonal permutations of individual welfare functions. Hence Theorem 2.[4]

Friedman's criticism of Lerner, quoted above, treats him as a prisoner of utilitarianism, but Lerner is free to leave the Benthamite jail.

[3] It can, however, be easily weakened.

[4] Note that for this result even Assumption 2 is much too strong and the quasi concavity of W is sufficient.

REFERENCES

P. Dasgupta, A. K. Sen, and D. Starrett, "Notes on the Measurement of Inequality," *J. Econ. Theor.*, Apr. 1973, *6*, 180–87.

M. Friedman, "Lerner on the Economics of Control," *J. Polit. Econ.*, Oct. 1947, *55*, 405–16; reprinted in M. Friedman, *Essays in Positive Economics*, Chicago 1966.

A. P. Lerner, *The Economics of Control*, New York 1944.

R. A. McCain, "Distributional Equality and Aggregate Utility: Further Comment," *Amer. Econ. Rev.*, June 1972, *62*, 497–500.

M. McManus, G. M. Walton, and R. B. Coffman, "Distributional Equality and Aggregate Utility: Further Comment," *Amer. Econ. Rev.*, June 1972, *62*, 489–96.

M. Rothschild and J. E. Stiglitz, "Some Further Results on the Measurement of Inequality," *J. Econ. Theor.*, Apr. 1973, *6*, 188–204.

A. K. Sen, "Planners' Preferences: Optimality, Distribution and Social Welfare," in J. Margolis and H. Guitton, eds., *Public Economics*, London, New York 1969.

———, *Collective Choice and Social Welfare*, San Francisco 1970.

———, *On Economic Inequality*, Oxford, New York, forthcoming.

Specification Error In Macro-Econometric Models: The Influence of Policy Goals

By JAMES R. CROTTY*

Econometric estimation of macro-economic relations has typically been based on the assumption that the monetary and fiscal instruments of the government are exogenous variables whose movements are not predicated in any systematic way on changes in economic conditions. Yet it has been recognized for some time that this assumption is in fact incorrect, and that any estimation procedure which relies on this assumption may be subject to important specification error. John Wood, for example, has cautioned that "If government policy makers do not behave randomly with respect to economic events . . . " then the typical estimation procedure " . . . results in inconsistent estimates of the parameters in equations describing the behavior of the private sector of the economy" (p. 135). Michael Hamburger, pursuing Wood's lead, has argued for a new approach which recognizes " . . . that it is the ultimate objectives of . . . policy (e.g., the desired rate of inflation) that should be treated as exogenous rather than the instruments used to achieve the objectives" (p. 34).

The purposeful nature of economic policy has not only been recognized by most economists; in the work of Henri Theil and the Netherlands School we have been provided with a formal theory of optimal government economic policy.[1] It assumes that the government or state behaves as if it possesses a preference function which orders possible outcomes related to a set of economic instrument and goal variables. The goals are related to the instruments through a representation of the structure of the economy. The purpose of this paper is to use the Theil framework to demonstrate explicitly the serious nature of the specification error that may arise if the preference function underlying government policy is ignored in macroeconometric work. The analysis is developed in a reduced-form context, but the specification error discussed in the paper applies to structural estimation as well.

I. The Model

Following Theil, we make the following two assumptions:

1) All vectors $X=[X_h]$, $Y=[Y_i]$ of instruments and noncontrolled or goal variables, respectively, are real valued. Deviations $\overline{X}=[X_h-X_h^*]$ and $\overline{Y}=[Y_i-Y_i^*]$ of these variables from their desired values, X^* and Y^*, are ordered according to the state decision maker's preference such as to allow a representation by means of a quadratic disutility function:[2]

$$W(\overline{X},\overline{Y}) = \alpha'\overline{X} + \beta'\overline{Y} + 1/2\,(\overline{X}'A\overline{X} + \overline{Y}'B\overline{Y} + \overline{X}'C\overline{Y} + \overline{Y}'C\overline{X})$$

where $\alpha=[\alpha_h]$, $\beta=[\beta_i]$, $A=[a_{hk}]$, $B=[b_{ij}]$, $C=[C_{hi}]$ are vectors and matrices of fixed elements and appropriate order. For ease of exposition we will treat the C matrix as null, implying that the state does not associate specific tools with specific goals, and assume that A is diagonal, implying that the state is not concerned with the relative position of the instruments.

2) The vectors X, Y are connected by a

* Assistant professor of economics, Bucknell University. I am grateful to Michael Hamburger, Leonard Rapping, the managing editor, and especially Raford Boddy for helpful suggestions. The reader's attention is directed to a recent article by Stephen Goldfeld and Alan Blinder which deals with problems similar to those discussed in this paper but which appeared too late to be incorporated here.

[1] See Theil for an extensive discussion of the theory and applications of his approach.

[2] The term X^* refers to preferred levels of instruments abstracting from their impact on Y. For an explanation of the necessity of including the instruments in the state preference function, see Theil, p. 80, and Arthur Okun, pp. 128–34.

linear equation system $Y=RX+S$ where $R=[r_{ih}]$ is a matrix of fixed elements describing the multiplicative structure of the constraints related to instruments and $S=[s_i]$ is a vector of fixed elements describing the additive structure. The system $Y=RX+S$ is the reduced form of the economy, which is a constraint in the state's decision problem. We abstract from any uncertainty associated with the reduced form and treat the problem as one of certainty equivalence.

The model thus characterizes an optimal decision-making problem in which the state wishes to handle its policy instruments in such a way as to minimize the value of its quadratic disutility function subject to the constraints represented by the reduced form of the economy. In other words, the problem is to choose X in order to:

$$\text{(1)} \qquad \text{minimize } W=\alpha'\overline{X}+\beta'\overline{Y}+1/2\,[\overline{X}'A\overline{X}+\overline{Y}'B\overline{Y}]$$
$$\text{subject to } Y=RX+S$$

We can find the optimal value of the instruments, X^0, by substituting the constraint into the preference function, partially differentiating with respect to the vector X, and setting the resulting expression equal to the zero vector.

$$\text{(2)} \qquad \frac{\partial W}{\partial X}=(\alpha+R'\beta)+A\overline{X}+R'B\overline{Y}=0$$

or

$$\text{(3)} \quad \alpha+A(X-X^*)=-R'\beta-R'B(Y-Y^*)$$

These expressions represent the first-order conditions for an optimal choice, i.e., they must hold at $X=X^0$. They are not the linear decision rules generally associated with quadratic preference function problems. Linear decision rules or response functions relate X^0 to exogenous and lagged endogenous variables. In this case, the linear decision rules would be

$$X^0 = X^0(S,\ Y^*,\ X^*;\ \alpha,\ \beta,\ A,\ B,\ R)$$

The expressions in (3) are the marginal utility conditions characterizing an optimal policy. The left-hand side of the equations are estimates of the marginal changes in utility resulting from movements in X. The right-hand side are the marginal changes in utility which would come about through the resultant movements in Y. At X^0 they must balance

Since most macro-economic reduced-form regressions have Y as the dependent variable, we must rearrange the optimality conditions. For convenience, assume that R is a square matrix. That is, we assume that there are an equal number of goals and instruments. Note that this does not imply that all desired values can be achieved simultaneously.[3]

$$\text{(4)} \quad \overline{Y} = -(R'B)^{-1}(\alpha+R'\beta)-(R'B)^{-1}A\overline{X}$$

Because reduced-form regressions are generally not in deviation form, transform (4) as follows:

$$\text{(5)} \quad Y = -(R'B)^{-1}(\alpha+R'\beta)+Y^* + (R'B)^{-1}AX^*-(R'B)^{-1}AX$$

If we assume that Y^* and X^* do not change over time, i.e., that the policy problem is exclusively one of stabilization, we can write (5) as:

$$\text{(6)} \qquad Y=\delta-(R'B)^{-1}AX$$

where δ is a vector of constants.

The optimality condition, (6), is a relation between Y and X. It is *not* the reduced form. The reduced form, $Y=RX+S$, is a constraint on the state in this model. Both sets of relations, $Y=\delta-(R'B)^{-1}AX$ and $Y=RX+S$, must be satisfied. It will be shown below that, with respect to the role of R, the relation between Y and X in the optimality

[3] If the state is able to achieve all of its desired values for goals and instruments simultaneously, then the preference function parameters have no empirical significance and the reduced-form parameter estimates are free of the particular specification error which is the concern of this paper. For a discussion of this case, see Wood, pp. 157–60. However, the argument that, with respect to major economic goals and policy instruments, the economy has always been approximately where the state desired it to be might be rather embarrassing to the state policy makers.

condition may be different than—indeed even the opposite of—their relation in the constraint. Coefficients of a regression relating Y to X will generally be influenced by both (6) and the reduced form constraint. This situation is similar to the price-quantity relation in the demand and supply functions for a commodity. Price-quantity time-series data will tend to trace out whichever function is relatively more stable over the sample period. Under the assumptions of this model, the reduced form is a relation between Y and X, *given* S. The optimality condition is a relation between Y and X, *given* Y^* *and* X^*. Regression estimates will be dominated by whichever relation is relatively more stable over time: S clearly changes every period, shifting the reduced form $Y-X$ relation; Y^* and X^* may change, in which case the optimality condition $Y-X$ relation will shift. If the former shift is larger than the latter the optimality condition will dominate the regression estimates.

We will examine two static cases, i.e., cases in which the state in its planning is concerned only with the current period. Assume for both cases that the state has two goals, Y_1 and Y_2, and two instruments, X_1 and X_2. We are concerned with the interpretation of the estimated coefficients from a linear regression relating Y_1 as the independent variable to X_1 and X_2. The question at issue is: What can we validly infer from the relative size of the regression coefficients about the impact multipliers, r_{11} and r_{12}, of the reduced form? Since we are concerned in this paper with the direction and extent of the specification bias caused by neglecting to consider the goals of the state, we will assume that the regression coefficients are estimates of the optimality conditions. This assumption will enable us to judge the extent of the maximum specification bias. The maximum bias would occur, as suggested above, when the reduced-form $Y-X$ relation exhibited large shifts over time relative to the optimality condition $Y-X$ relation.

Case 1

Assume that B is a diagonal matrix. This implies that the cardinal disutility measure associated with a squared deviation of Y_1 from Y_1^* does not depend on the level of $\overline{Y}_2$, and conversely. With this specification, the $\overline{Y}_1$ regression would yield estimates of the following relation:

$$(7)\qquad Y_1 = \delta_1 - \frac{a_{11}b_{22}r_{22}}{b_{11}b_{22}(r_{11}r_{22} - r_{21}r_{12})} X_1 + \frac{a_{22}b_{22}r_{21}}{b_{11}b_{22}(r_{11}r_{22} - r_{21}r_{12})} X_2$$

Given the knowledge that the regression coefficient of X_1 is larger in absolute value than that of X_2, what can we validly infer about the relative absolute values of r_{11} and r_{12}? In this case, the answer is almost nothing! A large relative regression coefficient for X_1 would imply either:

1) a_{11} is larger than a_{22}. This means that the state is more concerned about moving X_1 away from its desired value than it is about moving X_2.

And/or

2) r_{22} is larger than r_{21}. This implies that the impact on Y_2 of X_2 is more powerful than that of X_1.

Case 2

Assume that B is not a diagonal matrix so that the utility associated with one goal depends upon the condition of the other goal. This yields the following relation:

$$(8)\qquad Y_1 = \delta_1 - \frac{a_{11}(r_{12}b_{12} + r_{22}b_{22})}{N} X_1 + \frac{a_{22}(r_{11}b_{12} + r_{21}b_{22})}{N} X_2$$

where $N = (b_{11}b_{22} - b_{21}b_{12})(r_{11}r_{22} - r_{21}r_{12})$

As with Case 1, we cannot infer from larger regression coefficients for X_1 than X_2 anything about the relative size of r_{11} and r_{12} unless we know the relative sizes of the other parameters. This case does introduce a particularly perverse condition however. Note that r_{12} appears in the numerator of the X_1 coefficient while r_{11} appears in the numerator of the X_2 coefficient. Therefore, under many plausible assumptions about the size of the

parameters other than r_{11} and r_{12} we would find that a relatively large regression coefficient for X_1 as opposed to X_2 would indicate that X_2 was really more powerful than X_1.

II. The Role of Lagged Instruments

Reduced-form regressions frequently include lagged instrument values as independent variables. There are several ways to generalize the preference function framework in order to show that current values of the goals may be related to both current and lagged values of the instruments in such a way that the regression coefficients of lagged instruments contain both reduced-form and preference function parameters. For example, consider a static state preference function in which the disutility associated with current $\overline{Y}_t$'s depends upon their lagged values. Such a function would represent a situation in which the state's concern with an unemployment rate which was 1/2 percent above the desired value would be different if the unemployment rate was rising than if it was falling. One example of this class of preference functions would be:

$$(9)\quad W(\overline{X}_t, \overline{Y}_t) = \alpha'\overline{X}_t + \beta'\overline{Y}_t + 1/2\,[\overline{X}_t' A\overline{X}_t + \overline{Y}_t' B\overline{Y}_t + \overline{Y}_t' D\overline{Y}_{t-1}]$$

where the subscripts refer to time periods.

The optimal choice conditions for this function are:

$$(10)\quad \frac{\partial W}{\partial X_t} = (\alpha + R'\beta) + A\overline{X}_t + R'B\overline{Y}_t + R'D\overline{Y}_{t-1} = 0$$

or, under the assumption that R is a square matrix:

$$(11)\quad \overline{Y}_t = -(R'B)^{-1}(\alpha + R'\beta) - (R'B)^{-1}A\overline{X}_t - (R'B)^{-1}D\overline{Y}_{t-1}$$

Since $Y_{t-1} = RX_{t-1} + S_{t-1}$ we have:

$$(12)\quad \overline{Y}_t = -(R'B)^{-1}(\alpha + R'\beta) - (R'B)^{-1}A\overline{X}_t - (R'B)^{-1}D(RX_{t-1} + S_{t-1} - Y^*_{t-1})$$

In general, S_{t-1} will contain distributed lagged values of X. Therefore, (12) will relate $\overline{Y}_t$ to current and lagged values of X. Further the coefficients of both current and lagged values of the instruments will contain both preference function and reduced-form parameters.

III. Some Examples

Although the main purpose of the paper is to focus on the problem of specification error by presenting a general analytical framework, a few examples may help to make the basic implications more evident. The examples which follow are not presented as definitive explanations of actual empirical results. Rather they are designed to emphasize the importance of preference function considerations in the interpretation of macro-econometric analysis.

A specific example of macro-economic reduced-form estimation that largely mirrors the assumptions made above is the evidence presented by Leonall Andersen and Jerry Jordan in support of the St. Louis equation.[4] Andersen and Jordan assess the strength and reliability of monetary and fiscal policy through the use of regressions relating changes in nominal GNP to current and lagged values of frequently used indicators of monetary and fiscal action. They assume that this equation is a reduced form which is " . . . a derivable consequence of a system of equations which may be hypothesized to represent the structure of the economy (i.e., a so-called structural model)" (p. 16). On the basis of this assumption and their regression results they conclude that "The response of economic activity to monetary actions compared with that of fiscal actions is (I′) larger, (II′) more predictable and (III′) faster. . . . All of these implications point to the advisability of greater reliance being placed on monetary actions rather than on fiscal actions" (p. 22).

If we assume that the state is concerned with one other goal in addition to GNP, price stability for example, then the Andersen-

[4] For other tests of the St. Louis equation see Edward Corrigan, Richard Davis, Frank de Leeuw and John Kalchbrenner, Arthur Laffer and R. David Ranson, and William Silber.

Jordan regression equation is identical in form to equation (12). If we neglect lagged terms it is identical in form to equations (7) and (8). Under the assumptions of *Case 1*, the regression coefficients of the monetary and fiscal instruments tell us nothing about the relative size of their *GNP* multipliers, i.e., about the relative strength of monetary and fiscal policy. Under the assumptions of *Case 2*, a relatively large coefficient for the monetary as opposed to the fiscal instrument could, as explained above, indicate that the fiscal instrument was more powerful than the monetary instrument—a conclusion diametrically opposed to the one drawn by Andersen and Jordan from the same empirical evidence.

With respect to policy oriented reduced-form regressions in general, different assumptions about the role of instruments and goals in the state's preference function will lead to different expectations regarding the sign and relative size of regression coefficients. Suppose that the state believed that the existence of a Phillips' curve relationship made it unnecessary to use more than one instrument to simultaneously influence aggregate demand and the price level and attached no cost or disutility to movements of that instrument away from its desired value. The optimal use of the instrument for stabilization purposes would lead to the failure of standard reduced-form regressions to establish any relation between the instrument and the goals. Instrument movements would exactly counterbalance other forces tending to move the goal variables away from their desired values.[5] Therefore, if a fiscal variable was the sole or chief stabilizing instrument, we would expect it to have a regression coefficient not significantly different from zero. Under the assumption of quadratic instrument adjustment costs both instruments would be used. We would expect to observe correlations between goals and instruments which were opposite in sign to the respective reduced-form multipliers since the disutility attached to instrument adjustment would prevent complete goal variable stabilization.[6]

As a final example, assume that *GNP* and the interest rate are the two goal variables. The interest rate might be included either because it was considered to be a goal variable in and of itself or, more likely, because it served as an intermediate operating policy target variable. Any change in demand or expenditure variables which tended to push *GNP* above its desired level would also tend to raise the level of the interest rate above its desired value. If, in the face of such demand pressure, the state reacted by increasing the money supply or monetary base in an attempt to keep the interest rate stable we would observe a positive correlation between *GNP* and the monetary instrument.[7] This possibility is most likely to occur if the monetary authority is associated in a special way with the interest rate target. Since central banks do seem to place important emphasis on financial market and interest rate stability, this line of reasoning may offer a partial explanation of the fact that reduced-form regressions of the St. Louis variety consistently establish positive coefficients for the money supply or monetary base.

IV. Conclusion

The fundamental conclusion of this paper is that if the government is using its policy instruments in a systematic attempt to influence economic activity, there may be substantial specification error in any macroeconomic empirical work which does not utilize policy response functions. The policy decision-making process can generate relation-

[5] John Kareken and Robert Solow, p. 16, have pointed out this pitfall in the interpretation of regressions relating policy goals and instruments.

[6] See Brainard for a case in which incomplete goal variable stabilization results from the uncertainty attached to the policy maker's estimates of the reduced-form multipliers.

[7] This explanation of a positive money supply regression coefficient was suggested by Hamburger. Okun, p. 151, discusses the possibility that Federal Reserve policy may produce a positive relationship between the money supply and *GNP* by accommodating changes in the demand for money which arise from shifts in aggregate demand for goods and services. A related explanation by Kochin is that the monetary authorities consistently overreact to changes in aggregate demand by adjusting the money supply by more than is necessary to achieve stabilization.

ships between variables which are unrelated to or even inversely related to the true parameters of the economic structure or model. The specification error is potentially quite severe. Therefore, the parameters of the economic model will have to be estimated jointly with those of the state preference function.

REFERENCES

L. Andersen and J. Jordan, "Monetary and Fiscal Actions: A Test of Their Relative Importance in Economic Stabilization," *Fed. Reserve Bank St. Louis Rev.*, Nov. 1968, *50*, 11–24.

W. Brainard, "Uncertainty and the Effectiveness of Policy," *Amer. Econ. Rev. Proc.*, May 1967, *57*, 411–25.

E. Corrigan, "The Measurement and Importance of Fiscal Policy Changes," *Fed. Reserve Bank New York Mon. Rev.*, June 1970, *52*, 133–45.

R. Davis, "How Much Does Money Matter? A Look at Some Recent Evidence," *Fed. Reserve Bank New York Mon. Rev.*, June 1969, *51*, 119–31.

F. de Leeuw and J. Kalchbrenner, "Monetary and Fiscal Actions: A Test of Their Relative Importance in Economic Stabilization-Comment," *Fed. Reserve Bank St. Louis Rev.*, Apr. 1969, *51*, 6–11.

S. Goldfeld and A. Blinder, "Some Implications of Endogenous Stabilization Policy," *Brookings Papers*, Washington 1972, *3*, 585–640.

M. Hamburger, "Indicators of Monetary Policy: The Arguments and the Evidence," *Amer. Econ. Rev. Proc.*, May 1970, *60*, 32–39.

J. Kareken and R. Solow, "Lags in Fiscal and Monetary Policy," in *Stabilization Policies, Commission on Money and Credit*, Englewood Cliffs 1963.

L. Kochin, "Judging Stabilization Policy," unpublished doctoral dissertation, Univ. Chicago 1973.

A. Laffer and R. Ranson, "A Formal Model of the Economy," *J. Bus., Univ. Chicago*, July 1971, *44*, 247–70.

A. Okun, "Fiscal-Monetary Activism: Some Analytical Issues," *Brookings Papers*, Washington 1972, *1*, 123–63.

W. Silber, "The St. Louis Equation: "Democratic" and "Republican" Versions and Other Experiments," *Rev. Econ. Statist.*, Nov. 1971, *53*, 362–67.

H. Theil, *Optimal Decision Rules for Government and Industry*, Amsterdam 1967.

J. Wood, "A Model of Federal Reserve Behavior," in G. Horwich ed., *Monetary Process and Policy: A Symposium*, Homewood 1967.

Import Controls on Foreign Oil: Comment

By Ryan C. Amacher, Robert D. Tollison, and Thomas D. Willett*

The question of whether controls on the importation of foreign oil into the United States should take the form of tariffs or quotas has been a topic of recent public debate and investigation by economists. Under static competitive conditions it is well known that equivalent tariffs and quotas can be constructed. Hence in this context, there is no choice to be made on economic efficiency grounds.[1] However, in a recent issue of this *Review*, George Hay points out that the actual market for oil in the United States differs from the required textbook conditions for equivalence. Under the *U.S.* oil import program which prevailed until recently, each refiner's quota for import of foreign oil is a positive function of his refinery input. Since import tickets are allocated free of charge, rather than auctioned, the form of the quota lowers the marginal cost of domestic refiners. Hay goes on to show that when combined with other static competitive assumptions, this quota mechanism could generate greater consumer benefits in terms of lower prices than would an equivalent tariff (equivalent in the sense that the same percentage of imports is admitted).[2]

Hay expresses a preference for tariffs in a real world context and warns that his analysis of the price effects of the *U.S.* oil quota system holds only under very restrictive conditions. However, he does not address what is perhaps an even more important deviation of the domestic oil industry from the standard textbook model: crude oil production in the United States was limited in the major producing states by regulatory commissions that practiced market demand prorationing under the old oil quota program. Under this system, any price set by the industry is ratified by the commissions by limiting production to a level that will not result in the accumulation of undesired inventories.[3]

We are not addressing the issue of the level of price in the oil industry in this paper. Rather, we wish to review the effects of tariffs and quotas on resource allocation, an issue which Hay omits from his analysis; and, for this purpose we make use of the simple model of a profit-maximizing monopoly as a characterization of the domestic oil industry. The assumption of profit maximiza-

* University of Oklahoma; U.S. Council of Economic Advisers and Texas A&M University; and U.S. Treasury; respectively. Tollison and Willett were at Cornell University during the original drafting of the paper. Without necessarily implying agreement with all parts of our analysis, we should like to thank George Borts, George Hay, Edward Mitchell, Edward Tower, and an unknown referee for helpful comments on an earlier draft and the Bureau for Business and Economic Research, University of Oklahoma, for drafting services.

[1] See Jagdish Bhagwati (1955, 1968), Hirofumi Shibata, and Tsvi Ophir. As has been noted by a number of economists, on economic efficiency grounds neither a tariff nor a quota is probably the first best method of correcting for the national defense externalities associated with the oil market.

[2] Hay assumes the domestic supply of crude oil can be taken to be essentially unlimited at a constant price. Assuming that the prices of both foreign and domestic crude oil are fixed, it is easy to show that a quota system that allocates import tickets as a subsidy to refinery inputs results in lower consumer prices of refined products than a tariff which collects the foreign-domestic price differential for the Federal Treasury. This conclusion, as Hay points out, does not arise from any geographic considerations. If all domestic production and refining were at one point, it would still hold. Except in unusual circumstances all crude oil imported to Petroleum Administration for Defense (*PAD*) districts I–IV is used by refineries on the East Coast, the end point in Hay's analysis.

[3] This process of industry pricing is consistent with a number of models of price level formation in the industry. For example, it is consistent with a model stressing that colluding firms (in this case, the prorationing authorities) may limit price in order to forestall government entry through import or antitrust action. Hay notes the existence of prorationing, but uses it as an explanation for his assumption that the offer curve of domestic producers (at the wellhead) is horizontal for broad segments (1971, p. 688). He appears to have in mind a description of the industry that views the regulatory agency as attempting to maintain constant prices for given periods of time. For further discussion of the structure of the oil industry, see the Cabinet Task Force on Oil Import Controls, especially pp. 81 and 121.

tion is sufficient to show the differential effects of a tariff and quota on resource allocation in the industry. In this formulation the level of domestic reserves, and hence production, depends upon the price when the domestic industry is viewed as a profit-maximizing monopoly setting prices. As is well known under conditions of potential or actual domestic monopoly, the equivalence between quotas and tariffs breaks down.[4] The relevance of this point to the oil import question was stressed by the report of the Cabinet Task Force on Oil Import Control. As is illustrated below, where the foreign elasticity of supply to the domestic market is high (as it is in the case of oil), a nonprohibitive tariff still allows imports to overcome the misallocational effects of the domestic monopoly, while under a quota this would not be the case.

I. A Graphical Analysis

Figure 1 represents the market derived demand curve in the United States for crude oil. The colluding firms in this case behave like a textbook monopolist, setting marginal revenue (MR_m) equal to marginal cost (MC_m) with a resulting price of P_1 and output per unit of time of A. It can be shown

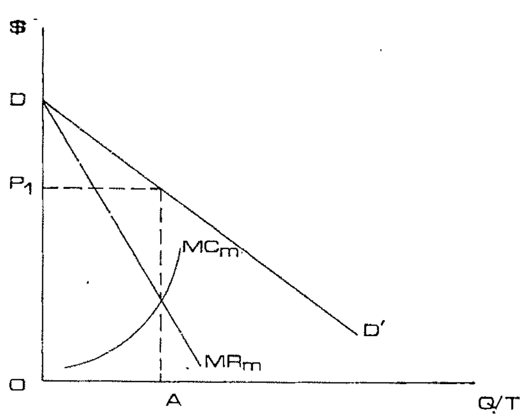


FIGURE 1: THE MONOPOLY WITH IMPORTS PROHIBITED

[4] See Bhagwati (1965) and Charles Kindleberger, pp. 566–68. We are concerned here with static comparisons. For another major condition for nonequivalence, caused by moving from a static to a dynamic model, see Mordechai Kreinin and Ingo Walter.

that P_1 will exceed the world price. Moreover, the domestic monopoly with imports prohibited results in the expected allocative inefficiency due to monopoly where marginal cost is less than price.

If the government introduced foreign sources of supply through a quota system, the market demand curve facing the domestic monopolist would shift to the left, as in Figure 2, by the amount of the quota ($C'D'$). The colluding firms would thus further restrict their output, maximizing short-run profit at price OP_2 and output OB.[5] The effect of the quota (as measured, for example, by the divergence of price and marginal cost in Figures 1 and 2) reduces but does not eliminate the market power of the domestic monopoly since marginal cost is still less than price.

If, instead of introducing the foreign supply of oil through quotas, the government had chosen an equivalent tariff (equivalent in that it established the same discrepancy between foreign and domestic price), the market power of the dominant domestic firms would have been severely constrained.[6]

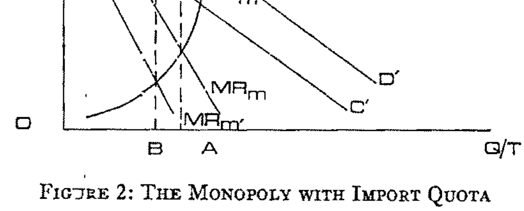


FIGURE 2: THE MONOPOLY WITH IMPORT QUOTA

[5] With a straight line demand curve (as drawn in Figure 2) there is an unambiguous shift in the monopolist's marginal revenue, indicating a decline in production. Under other assumptions (e.g., if the demand curve representing the domestic market is convex from below) it is uncertain whether the marginal revenue (and hence production) shifts up or down in response to an increase in the quota.

[6] Bhagwati demonstrated that in the presence of domestic monopoly a quota produces a higher price

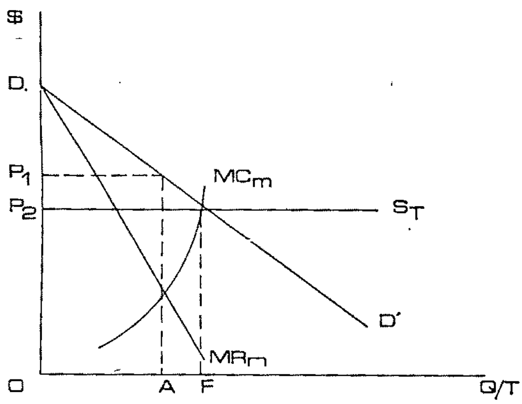


FIGURE 3: THE MONOPOLY WITH IMPORT TARIFF

In Figure 3 a tariff that establishes price OP_2 is levied. P_2S_T now represents the supply curve to the *U.S.* market. The once dominant coalition of firms, now without market power, produces output OF where price (OP_2) equals marginal cost. Market power is absent under the import tariff in the sense that in equilibrium the domestic price (the world price plus tariff) for crude no longer exceeds marginal cost.

The tariff and quota thus have quite different effects on the structure of the domestic oil industry. A quota allowed the domestic monopoly to maintain its power, since prorationing and excess capacity can continue to exist. The tariff, on the other hand, allows competitive supply at the world (plus tariff) price.[7] Under the tariff the domestic firms are forced to produce OF to the right of OA, while with the quota they are allowed to restrict output, producing OB to the left of OA.[8]

Switching to a tariff served to generate a better allocative solution by curbing domestic monopoly power and by stimulating an increase in domestic production. We suspect this allocative gain would considerably outweigh the type of potential price gains for consumers from a quota system as analyzed by Hay.[9]

(implicit tariff) than does a tariff that allows the same volume of imports (1965, p. 53). Since we are concerned with resource allocation comparisons and not price comparisons (an issue on which Hay's analysis is vulnerable in the case of domestic monopoly, see E. Mitchell et al.), we consider here a tariff that allows the correct volume of imports to establish an equivalent domestic price. This is consistent with the standard definition of equivalence.

[7] See Kindleberger, p. 134. While Hay argues the possible case for a quota mechanism in terms of favorable price effects for consumers, and while he mentions some reasons that tariffs may be preferred to quotas, p. 691, he does not address the basic efficiency issue involved in choosing between the two policies.

[8] Inefficient producers would likely oppose tariff legislation to replace the quota, because the tariff has the effect of rationalizing production, bringing about the close of the inefficient well.

[9] In an unpublished paper by Mitchell et al., the issue of how final consumer prices are affected by a series of tariff and quota mechanisms in the presence of a domestic monopolist is discussed. This analysis demonstrates that under conditions of domestic monopoly the quota mechanism discussed by Hay turns out to generate the *highest* consumer prices. For a related analysis, see Rachael McCulloch and Harry Johnson.

REFERENCES

J. Bhagwati, "More on the Equivalence of Tariffs and Quotas," *Amer. Econ. Rev.*, Mar. 1968, *58*, 142–46.

———, "On the Equivalence of Tariffs and Quotas," in R. E. Baldwin et al., eds., *Trade Growth and the Balance of Payments: Essays in Honor of Gottfried Haberler*, Chicago 1965.

G. A. Hay, "Import Controls on Foreign Oil: Tariff or Quota?," *Amer. Econ. Rev.*, Sept. 1971, *61*, 688–91.

C. Kindleberger, *International Economics*, 4th ed., Homewood 1968.

M. E. Kreinin, "More on the Equivalence of Tariffs and Quotas," *Kyklos*, Fasc. 1, 1970, 75–79.

R. McCulloch and H. G. Johnson, "A Note on Proportionally Distributed Quotas," *Amer. Econ. Rev.*, Sept. 1973, *63*, 726–32.

E. Mitchell, R. Tollison, E. Tower, and T. D. Willett, "The Price Effects of Tariffs and Quotas Under Domestic Monopoly," unpublished manuscript, Cornell Univ. 1972.

T. Ophir, "The Interaction of Tariffs and Quotas," *Amer. Econ. Rev.*, Dec. 1969, *59*, 1002–05.

F. M. Scherer, *Industrial Market Structure and Economic Performance,* Chicago 1970.

H. Shibata, "A Note on the Equivalence of Tariffs and Quotas," *Amer. Econ. Rev.,* Mar. 1968, *58,* 137–42.

I. Walter, "On the Equivalence of Tariffs and Quotas: A Comment," *Kyklos,* Fasc. 1, 1971, 111–12.

The Cabinet Task Force on Oil Import Controls, *The Oil Import Question,* 1970.

Import Controls on Foreign Oil: Reply

By George A. Hay*

The position of Ryan Amacher, Robert Tollison, and Thomas Willett (ATW) can be summarized as follows: Assuming that the domestic industry behaves like a profit-maximizing monopoly, if the quota on imported crude oil is replaced by a tariff that is by assumption set so as to achieve the same domestic price of crude oil, the market power of the dominant domestic firms is "severely constrained."

In response I would contend that 1) their assumption about the level of the tariff differs from what I had assumed; 2) their conclusion about the market power of the dominant firms is subject to question; and 3) their assumption about the behavior of the domestic industry is difficult to reconcile with the available data.

The basic issue that I attempted to analyze in the original paper was whether the price of refined oil would be affected, if the quota is replaced by a tariff set so as to permit the same percentage of crude oil imports. ATW, in attempting to rebut my analysis, start by assuming that the tariff will be chosen to lead to the *same price* for domestic crude as a quota system. Had they instead worked with my original assumption, their notion of treating the domestic industry as a monopoly would have led to a more direct comparison with my own conclusion. Still in the monopoly model it is *not* obvious that prorationing will be discontinued. For under the assumption that the tariff will be chosen so as to lead to the same percentage of imports as the quota, the tariff will be raised to ratify any output restriction desired by the domestic industry.[1]

In the light of ATW's assumption that crude oil prices will be the same under either system, their conclusion that under a tariff the market power of the dominant domestic firms is severely constrained is questionable. It is of course true that under the ATW assumption the tariff has the advantage that the real economic cost to this country of acquiring the oil is reduced,[2] with the savings going to domestic producers as producers' surplus, but one would not normally characterize that result as a diminution of market power.[3]

Indeed, judging by the ATW diagrams, the producing industry should *prefer* the tariff, as profits would appear to be substantially higher in a static situation with the "equivalent" tariff. In view of what most observers would regard as a hostile industry position[4] towards the Task Force recommendation of replacing the quota by a tariff, one suspects that ATW have not captured the essence of the situation. My own view, as stated in the original paper, is that the real advantage of tariffs is that, to the extent they are not easily revised upward or are revised only with a lag, the ability of domestic producers to increase price above the *current* level as demand and supply conditions change is sharply curtailed. This indeed represents a diminution of the domestic industry's market power, but one which cannot be conveniently analyzed in a single-period model as ATW have tried to do.

Finally, ATW "make use of the simple model of a profit-maximizing monopoly as a characterization of the domestic oil industry" (p. 1031). My assumption had been that " . . . state regulatory commissions in the large producing areas control output to keep price constant in the face of month-to-month

* Associate professor of economics, Yale University.

[1] This is developed in the Mitchell, Tollison, Tower and Willett manuscript cited by ATW.

[2] Assuming that domestic producers' cost curves accurately reflect the complete social cost of extracting the oil.

[3] The point is that the base price for determining the level of the tariff is the current domestic price. Thus until and unless cost conditions change, the effective domestic price continues to be the monopoly price.

[4] Not limited in my view to what ATW characterize as inefficient producers.

TABLE 1

Year	Average Wellhead Price of Crude Oil (Dollars/Barrel)	Production of Crude Oil (Million Barrels)
1959	$2.90	2575
1960	2.88	2575
1961	2.89	2622
1962	2.90	2676
1963	2.89	2753
1964	2.88	2787
1965	2.86	2849
1966	2.88	3028
1967	2.92	3216
1968	2.94	3330

Source: James Burrows and Thomas Domencich.

shifts in demand" (p. 688). This was not to suggest that the price of crude is in any sense a competitive price; because of prorationing, prices are surely above costs. But for purposes of predicting the response to the substitution of a tariff for the quota, the assumption that the industry will act like a profit-maximizing monopolist may be unrealistic.

An examination of the relevant data proves interesting in this regard. Table 1 lists for the ten-year period preceding the Cabinet Task Force study the yearly average price of crude oil and the corresponding yearly production figures.[5] While it would not be impossible to construct a monopoly model to generate such numbers, my original assumption seems to fit the data sufficiently well to recommend it for the purposes I had in mind.

[5] The price of crude has dramatically increased since 1968, but it is also true that prorationing has been suspended during most of that period. Of course, it must also be noted that import controls on crude were more-or-less abandoned early in 1973.

REFERENCES

J. C. Burrows and T. A. Domencich, *An Analysis of the United States Oil Import Quota*, Lexington 1970.

G. A. Hay, "Import Controls on Foreign Oil: Tariff or Quota?," *Amer. Econ. Rev.*, Sept. 1971, *61*, 688–91.

E. Mitchell et al., "More on Tariffs Versus Quotas to Control Oil Imports," unpublished, Cornell Univ. 1972.

In Search of the "Wheel of Wealth": On the Origins of Frank Knight's Circular-Flow Diagram

By Don Patinkin*

One of the memories that every former student of Frank Knight undoubtedly carries with him is that of the circular-flow diagram (reproduced on p. 791 of this issue) which Knight used to illustrate the workings of the economic system. To the best of my knowledge, the first appearance of this diagram in Knight's writings was in the mimeographed material that he prepared for teaching purposes at the University of Iowa in the early 1920's.[1] Knight first published the diagram, however, in the four chapters on the economic system that he wrote for the volume of *Readings* prepared for the famous undergraduate Social Science Survey Course at the University of Chicago (Knight 1932). Subsequently, Knight reissued these chapters as a separate booklet entitled *The Economic Organization* (1933), in which form it became a standard feature of his graduate theory course at Chicago over the years. So it was only natural for me to dwell upon this diagram in my memoir on Frank Knight as a teacher which appears elsewhere in this issue.

I had always assumed that this diagram was original to Knight. Thus it came to me as a surprise to notice—in the process of writing this memoir—that Knight himself referred to it as "the familiar figure of the 'wheel of wealth' " (Knight 1933, pp. 60–61). My curiosity aroused, I began a search for the presumably earlier uses of this figure that rendered it so "familiar"—only to discover that the question is not so simple.

Let me first note that though this diagram is concerned with the circular flow of money, Knight used it not in the context of monetary theory, but in order to illustrate certain basic aspects of value theory: namely, the differentiated roles that families and businesses fulfill in the specialization and exchange that characterize a market economy; the ways in which the respective activities of these units are guided by the prices of productive services, on the one hand, and the prices of final goods and services, on the other; and the ways in which the interaction of these two sets of prices enables the economic system to

* Professor of economics, The Hebrew University of Jerusalem. Without burdening him with any responsibility for the conclusions of this note, I would like to express my deepest appreciation and indebtedness to my colleague, Ephraim Kleiman, who encouraged me to write it, and who continued to be the source of many valuable suggestions and criticisms throughout its preparation—including the suggestion for its title! I am also indebted to Paul Samuelson for his stimulating comments on an earlier draft of this note; needless to say, he too does not bear responsibility for its conclusions. I would also like to thank my assistants, Akiva Offenbacher and Reuven Nutkis, for their invaluable help with the detailed examination of the literature on which this note is based—a task which they carried out with gratifying care, accuracy, and responsibility. And my sincerest thanks to Vera Jacobs for her efficient and conscientious typing of this note through its various drafts. Once again, I would like to thank George Stigler for making Knight's unpublished papers available to me, and Glen Gilchrist for providing xeroxes of requested materials from these papers. I am also grateful to Stanley Fischer and Friedrich A. Lutz for providing me with xeroxes of material that was unavailable here. I wish finally to thank the Central Research Fund of the Hebrew University and the Israel Academy of Sciences and Humanities for research grants that made this and other assistance possible.

[1] This is the material that Glen Gilchrist has catalogued under the title "*Economics*, 1920's" and filed in Boxes 8–9 of Knight's papers. Gilchrist has informed me that " . . . most of the material in these boxes is undated, but I have found the date 1926 or 'revised 1926' on a couple of these items as well as a phrase indicating it was used as supplementary material for 'Economics IC' " (personal letter dated April 24, 1973). On the other hand, the bibliography of Knight's works which appears at the beginning of *The Ethics of Competition* (1935, p. 15, bottom) adds the following comment to its listing of Knight (1932): "These four items are an abstract of material mimeographed for private circulation at the University of Iowa during the years 1922–25." I presume that this is the mimeographed material in Boxes 8 and 9, and have accordingly dated this material as 1922–26.

fulfill its basic functions of determining what to produce, how to produce, and how to distribute the product.

As the first step in the search for the sources of this diagram,[2] I examined the standard American textbooks of the period in order to see if in addition to describing the process of specialization and exchange—and the role played by money in facilitating this process—they had also made use of the term wheel of wealth or of the diagram that Knight associated with it. Such an examination of the respective textbooks of Francis A. Walker (1887), Richard T. Ely (1893, 1937), Edwin R. A. Seligman (1905), Irving Fisher (1910, 1912), Frank W. Taussig (rev. ed., 1915) and Fred M. Taylor (1911, 1925) showed that they had not. The same is true of Herbert J. Davenport (1897, 1908, 1913)—to whom, as I have noted in my memoir, Knight had a special affinity. Nor does the term or diagram appear in the textbook by Leverett S. Lyon (1923), which was used at the University of Chicago Business School in the early 1920's. Similarly, no instances of such a term or diagram are to be found in such contemporary British standard works as Alfred Marshall's *Principles* (first ed., 1890; eighth ed., 1920), J. Shield Nicholson (1893–1901), Philip H. Wicksteed (1910; rev. 1933), Edwin Cannan (rev. ed. 1916), and Sydney Chapman (new ed., 1917). Similarly, there is no reference to the term or diagram "wheel of wealth" in Palgrave's *Dictionary of Political Economy* (new ed., 1925). And what lends added significance to the absence of such a diagram from all of these books is the fact that most of them did make use of analytical diagrams in other contexts.

In view of the important influence of the Austrian school on Knight's thinking, I then turned to the Continental literature—only to find that there is no discussion or depiction of the wheel of wealth in the classic works of either Carl Menger (1871), Eugen v. Böhm-Bawerk (1888), or Friedrich v. Wieser (1889). The same is true of the earlier widely used textbook of that leading member of the old historical school, Wilhelm Roscher (13th ed., 1878). Nor does the "wheel" appear in the standard treatises of the turn of the century by Eugen von Philippovich (1897–99) and Adolph Wagner (1892–94, 1907–09), (see fn. 11 below). Similarly, the wheel does not seem to have been referred to in the teaching of economics in Germany in the 1920's.[3] In any event, it does not appear in Gustav Cassel's book (1923, 1932), which was a standard text on the Continent at that time, though Cassel does provide a verbal description of the exchange by consumers of their productive services for goods produced by firms, in what is essentially a barter economy (Cassel 1932, pp. 43–44).[4] I must, however, admit that less significance can be attached to the absence of the wheel-of-wealth diagram from this literature as compared to the Anglo-American in view of the fact that this Continental literature generally did not make use of analytical diagrams in any context.

This consistent failure to find earlier instances of the wheel-of-wealth diagram then raised the possibility that perhaps the search was being conducted in the wrong direction: that perhaps the antecedents of the wheel were to be found not in the economic literature that was concerned primarily with value theory (as is largely the case with the foregoing works), but with that concerned with monetary theory and practice—and with the circular flow of money in particular.

[2] As an aside, I might note that the use of diagrams to describe the exchange of commodities goes back to Aristotle (Theocharis 1961, p. 4). Aristotle, however, used his diagram essentially to illustrate the labor theory of value, and not the circular flow of money and commodities. I am indebted for this reference to my colleague Ephraim Kleiman.

[3] I base this statement on information received from a sample of two: A. L. Gaathon (Gruenbaum) of the Bank of Israel Research Department, who completed his Ph.D. degree at the University of Berlin in 1934; and Freidrich A. Lutz of the Swiss Institute of International Studies in Zurich, who completed his at the University of Tubingen in 1925. Gaathon has also informed me that he has not found any reference to the wheel of wealth in the textbooks by Adolph Weber, Wilhelm Lexis, and Othmar Spann, which were widely used in Germany during the first decades of this century.

[4] Knight (1921–22) wrote a very favorable review article of the first edition of Cassel's book, though without any reference to the point now under discussion.

Such a literature goes back at least to the seventeenth-century mercantilists who (as Jacob Viner has told us) often described the circulation of money as a medium of exchange by means of "... analogies, expecially with the circulation of blood, which William Harvey had discovered not long before" (Viner 1937, p. 37). Even more relevant to our present purpose is Quesnay's *Tableau Economique* (1972 (1758–59), pp. ii–iv),[5] in which the circulation of goods and money among different classes (though not those designated by Knight) is analyzed and illustrated (though not by means of a wheel, and not for the purposes that concerned Knight).[6]

Quite naturally, the circular flow of money was also one of the topics frequently discussed in the quantity-theory literature. Thus in his pioneering work on the quantity-theory equation, Simon Newcomb emphasized that to the "monetary circulation" there corresponded a reverse "industrial circulation" of "wealth or services"—and used the term "societary circulation" to designate the combination of these two opposite circulations (1886, pp. 318–19, 326). Furthermore, Newcomb made use of diagrams (one of which is reproduced in Figure 1a) to illustrate this "societary circulation—and explained that the "arrowheads" which showed the direction of the monetary circulation could "also be considered to represent the industrial circulation, the latter flowing in the opposite direction from that of the arrows, but along the same veins" (1886, pp. 318–19, 326). It is, however, clear from Figure 1a that Newcomb was not concerned with the flows between the basic functional units (families and businesses) that concerned Knight; nor did Newcomb make use of his diagrams to illustrate the organization of production and distribution by the price system.

Much the same statements can be made for the many diagrams (for example, Figure 1b) illustrating the exchange of goods for money under different circumstances that appear in an early Hebrew textbook on money by Vienna- and Berlin-trained Chaim Dov Hurewitz (1900, pp. 105, 112, 119, 150, 153, and 156), which has recently been rediscovered by Ephraim Kleiman (1973). And the same can also be said for the diagram (Figure 1c) showing the opposite circular flows of goods and money that appears in Knut Wicksell's *Lectures on Political Economy* (1901, vol. 1, p. 105; vol. 1, p. 64 of the English translation)—whose purpose, however, is essentially to show how the use of money as a medium of exchange makes the "double coincidence" unnecessary. And in both cases one gets the clear impression that the authors present their diagrams not as original contributions on their part, but as something which they presume to be familiar to their readers from earlier sources—though they do not explicitly indicate what these sources were.[7,8]

[5] See also the passages from the "Extract from the Royal Economic Maxims of M. de Sully" in the *Tableau*, pp. 1–3.

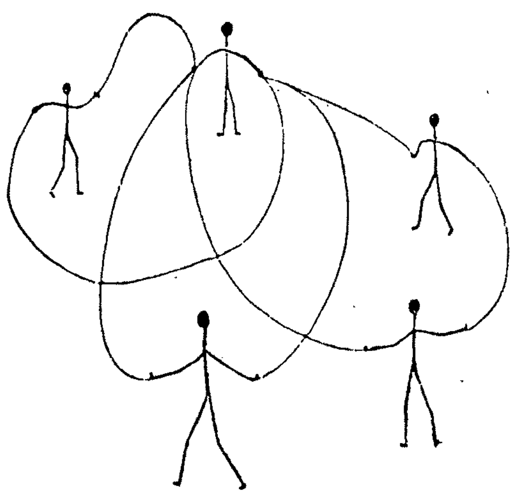
[6] My colleague David Levhari has reminded me of Karl Marx's detailed discussions in *Capital* of his M(oney) = C(ommodity) = M(oney) schema and variations thereof, which stem in part from Quesnay. Marx's various schemata, however, are quite different in both content and purpose from Knight's wheel of wealth. See Marx, v. 1, Part I, ch. 3, especially p. 125; v. 2, Part I, especially p. 113. See also Sweezy (1942, pp. 56–59, 138–45).

[7] In particular, as Kleiman (1973) has emphasized, "Hurewitz did not presume to present new theories, but rather wished to acquaint his readers with some part of what he considered to be the accepted corpus of economic theory." Still, as implausible as it may sound (and it *is* implausible—so that I would welcome counter-examples from the literature), Hurewitz's diagram is the first that I have been able to find that presents the opposite flows of goods and money in the by-now familiar "double-circle" form.

[8] The literature on the circular flow of goods and money has been surveyed by Arthur W. Marget from a somewhat different—and much broader!—viewpoint in the course of his discussion of the history of "The Income Approach to the Theory of Prices" (1938, vol. 1, ch. 12; see also 1942, vol. 2, ch. 7). See also Joseph A. Schumpeter (1912; 1954, ch. 2). Marget (1942, v. 2, pp. 356–57, fn. 18 and 19) also refers to circular-flow diagrams by Francesco Ferrara (1864; 1938, pp. 85, 86, 92) and Fleeming Jenkin (1887, p. 150). All of these diagrams, however, refer not to a money economy, but to the circular flow of goods in a barter economy as a result of barter exchanges of one good for another among the individuals of the economy. Correspondingly, these diagrams are not relevant to the present discussion. Nevertheless, I cannot resist reproducing here Jenkin's

Perhaps Knight, too, considered his diagram to be such a "natural" description of the economic system that he took its familiarity for granted.[9] Still, I would conjecture that Knight's use of his diagram was directly inspired by certain earlier diagrams—though not those of Newcomb and Wicksell which (as already emphasized) only described the role of money as a medium of exchange, and did not depict any of the functional relations that were Knight's main concern. Instead, I think the relevant earlier diagrams were those that appeared in the literature of the first decades of this century that attempted to explain depressions in terms of inadequacies of demand generated by interferences with the circuit flow of money—and that, because of this denial of Say's Law, grew up outside the mainstream of economic thought (though they ultimately led to the development of the money-flow analysis of Copeland (1952) and the flow-of-fund analysis of the Federal Reserve (1955)). Even more specifically, I would conjecture that the direct source of Knight's wheel of wealth was the diagram (Figure 2) showing the circuit flow of money among different functional units that had been used in the foregoing context first by William Foster alone in an article (1922, p. 463), and then by Foster together with his colleague Waddill Catchings in their then widely discussed books on *Money* (1923, p. 305) and *Profits* (1925, p. 255), the latter of which Knight reviewed (1926).[10] I should also note Foster and Catchings' related contention that " . . . money spent in the consumption of commodities is the force that moves all the wheels of industry" (1923,

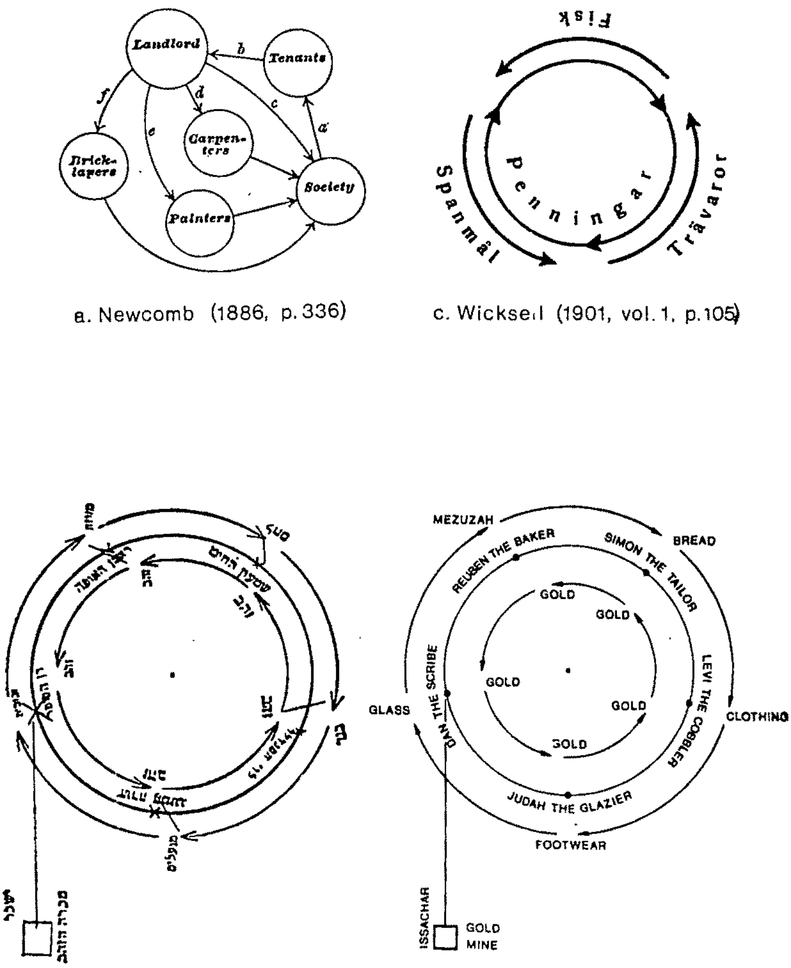
charming diagram illustrating the mutual benefits generated by the trade that takes place " . . . among five little rudimentary people standing upright," where "the lines going out of the strokes to the right indicate produce, which each has to sell . . . [and the] line arriving at his left hand represents the goods which each man consumes" (1887, p. 150–51).



I might also note that unlike Ferrara, Jenkin (who was a professor of electrical engineering) also discussed (though *sans* diagram) the process of trade in a money economy and in this context referred to the "closed circuit (as we might say, borrowing a metaphor from electrical science) . . . round which the sixpence travelled from consumer to producer while the goods went the other way from producer to consumer." (1887, p. 143). The familiarity of such a concept at the time is well-attested by Francis Ysidro Edgeworth's statement (in his biographical sketch of Jenkin (1899 1923, v. 2, p. 47)) that this metaphor "enhances some elementary principles of economics."

[9] Thus see Schumpeter's clear verbal description of some of the basic features of Knight's "wheel of wealth" in his [Schumpeter's] influential article on "Money and the Social Product" (1917/18; 1956, pp. 152–53). I might at this point also cite John R. Commons' observation (1934, p. 294) that " . . . from Quesnay to the Twentieth Century economic theory was dominated, in large part, by his analogy of Circulation of commodities and money. In the latter part of the Nineteenth Century it began to take over the analogy of Turnover. The one is the analogy of a 'flow,' the other of a 'wheel.' " It is not clear from this passage whether Commons actually attributed the use of the term "wheel" to the nineteenth-century literature itself—or whether (as I suspect) this represents Commons' own (and hence post-Knightian) use of the term. Unfortunately, Commons does not provide the references to the nineteenth-century literature that would enable us to check this question. In any event, to the best of my knowledge, Commons himself did not use this term in his earlier writings; see, for example, Commons (1893).

[10] The validity of specifically identifying the origin of Knight's wheel-of-wealth diagram with his review of Foster and Catchings' *Profits* depends, of course, on the assumption that the mimeographed material in which Knight first used this diagram was written in 1926—and I have above (see fn. 1) provided some evidence that this may indeed be the case. But even if Knight wrote this material earlier in the period 1922–26, it could still have been inspired by Foster and Catchings' earlier presentations of the circular-flow diagram—which as indicated in the text, began in 1922. On the intensive discussions—both lay and professional—generated at the time by Foster and Catchings' work, see Dorfman (1959, pp. 339–50).



a. Newcomb (1886, p. 336)

c. Wicksell (1901, vol. 1, p. 105)

b. Hurewitz (1900, p. 112)

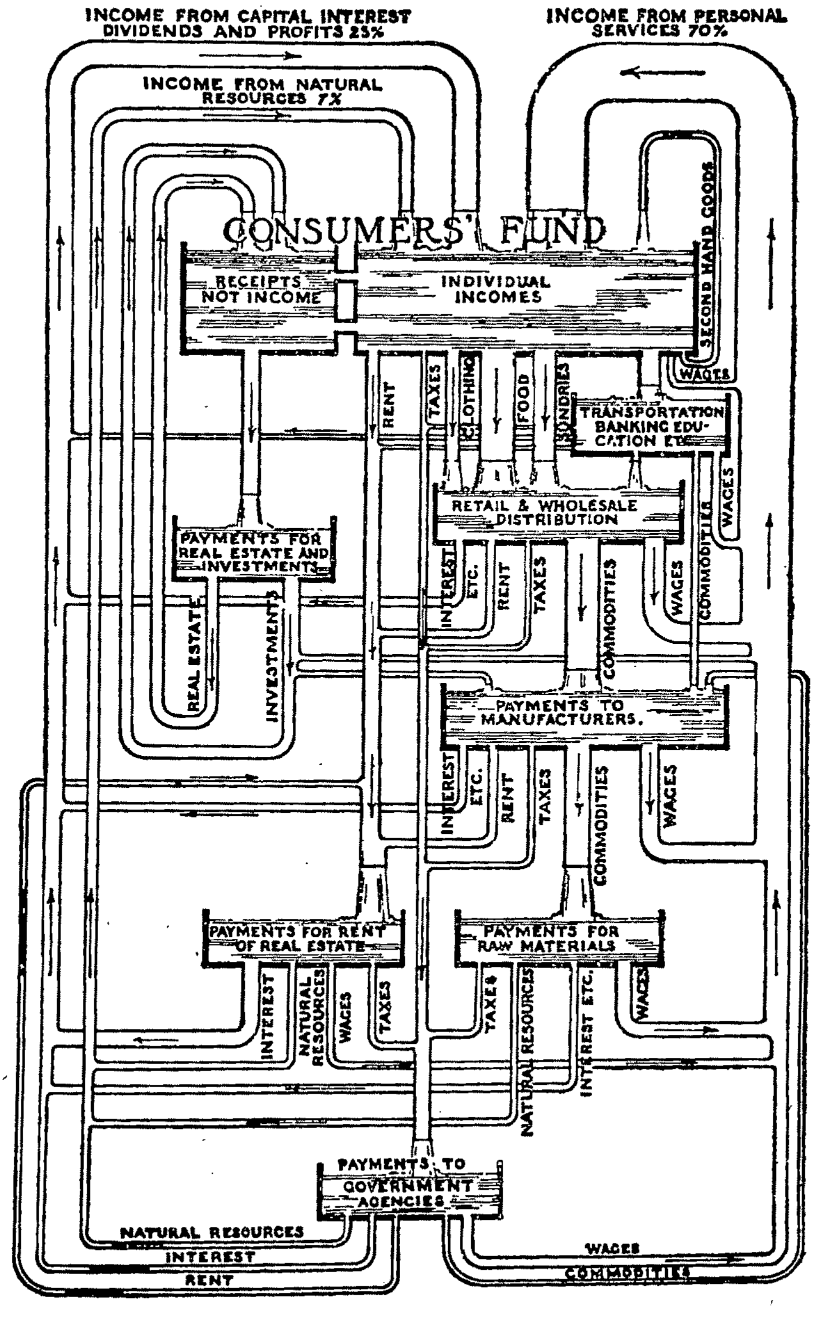
FIGURE 1

p. 277; 1925, p. 234) Thus Foster and Catchings not only presented a diagram showing the opposite flows of money and commodities between the functional units, households and firms, but also connected this diagram with "wheels"—albeit not those of wealth!

Foster and Catchings (1923, p. 303) cite as their source of the diagram the almost identical one by M. C. Rorty (1922, p. 63). Where did Rorty get it? He does not say. But I suspect that it can be traced back to the diagram (see Figure 3) which appears in the work of that "amateur American economist," Nicholas A. L. J. Johannsen, on *A Neglected Point in Connection with Crises* (1908, front matter)—and which has its origin in an earlier diagram by the same author (writing under the name J. J. O. Lahn) in a work entitled *Der Kreislauf des Geldes* (1903).[11] For though Johannsen was

[11] I am indebted to Friedrich Lutz for bringing "Lahn's" diagram to my attention. Marget (1938, vol. 1, pp. 331–36) discusses the intellectual influences that Johannsen and Adolph Wagner exerted one on the other. Whatever the nature of these influences, it is clear that Johannsen did not derive his diagram from Wagner: for, as noted above, Wagner made no use of diagrams in his analysis in general—and this is true of his analysis of money-flows in particular (1909, pp. 158 ff.).

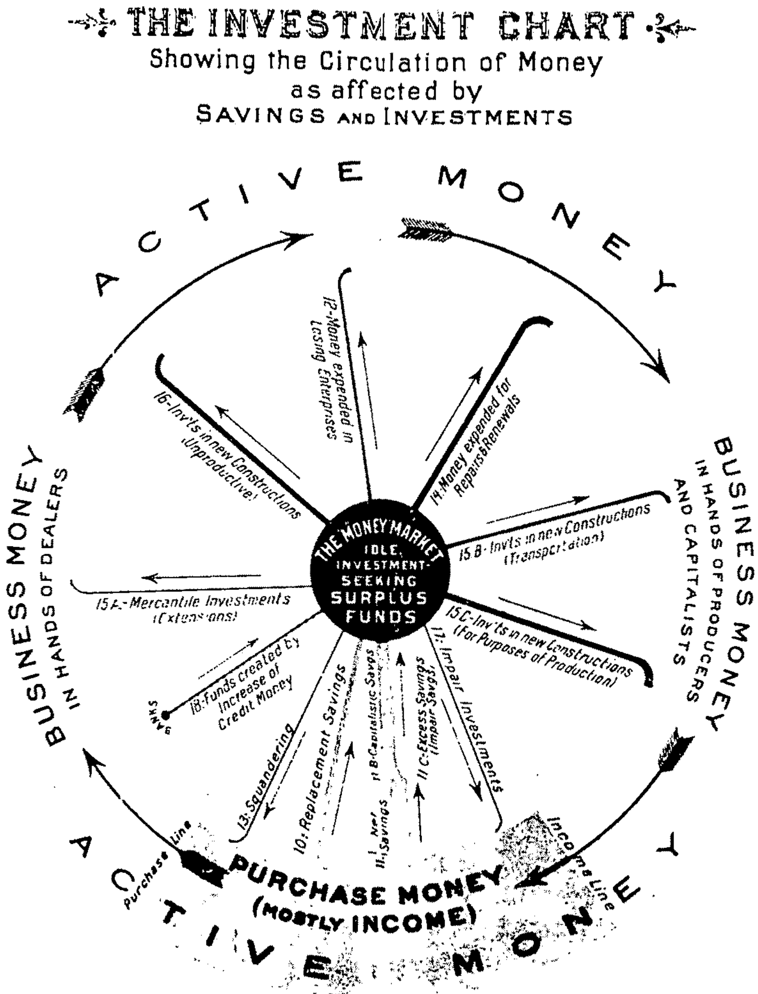
THE CIRCUIT FLOW OF MONEY



Reprinted by permission, Houghton-Mifflin Company.

Foster and Catchings' Diagram

Figure 2



Reprinted by permission, Augustus M. Kelly.

JOHANNSENS' DIAGRAM

FIGURE 3

neglected by most of the profession, his contributions were appreciated by Wesley C. Mitchell (Dorfman 1949, p. 413)[12]—who in 1920 became Director of the National Bureau of Economic Research, and to whom (among others) Rorty (who was then President of the Bureau)[13] expresses his general indebtedness.

Let me conclude with two comments. First, even if the source of Knight's "wheel of wealth" was Foster and Catchings' diagram, it must be emphasized that Knight adapted this diagram to completely different purposes. In particular, as indicated at the beginning of this note, Knight's concern was with the allocative functions of the price system under the implicit assumption of full employment, not with the possible generation of unemployment as a result of inadequacies of aggregate demand. Thus no matter what the origins of Knight's wheel of wealth as a diagram, its use to illustrate

[12] See also Mitchell's references to both Johannsen and Rorty as listed in the indices to his classic works on the business cycle (1913, 1927)

[13] See title page of Rorty's book.

basic aspects of price theory—as distinct from monetary theory—seems to have been original to him.[14]

Second, after having in this way built conjecture upon conjecture in the search for the "familiar figure" that inspired Knight's wheel-of-wealth diagram, I must—as something of an anticlimax—note the possibility that Knight never drew his inspiration from any such figure in the first place! For in the earlier mimeographed version (1922–1926) of his analysis of the functions of the price system that I referred to at the beginning of this note, Knight starts his discussion with the statement that "... the general character of an enterprise organization reduced to its very simplest terms can be indicated by a diagram which suggests the familiar figure of speech, the 'wheel of wealth'." Thus from this passage it would seem that the inspiration for Knight's wheel of wealth diagram originally came not from a "figure," but from a "figure of speech"—and that Knight actually developed his diagram independently of the earlier ones described above.

[14] In this connection it is interesting to note the different uses that Samuelson has made of circular-flow diagrams in his *Economics: An Introductory Analysis*—the textbook by means of which Knight's wheel of wealth has become familiar to a modern generation of economists. (On the origins of Samuelson's diagram in Knight's teachings, see above fn. 8.) All editions of Samuelson's textbook present a minor adaptation of Knight's description of the functions of the economic system in terms of "what, how, and for whom" to produce. But in the first three editions this discussion is not accompanied by Knight's wheel of wealth. Instead, Samuelson uses the circular-flow diagram in these editions only for the quite different purpose (to which Knight had also very briefly adverted (1933, p. 62)) of illustrating the equality between national income and national product (Samuelson 1948, p. 206; 1951, p. 231; and all subsequent editions up to the most recent—1970, p. 170). Only in the fourth (1958, p. 41) and later (e.g., 1970, p. 42) editions is Samuelson's discussion of the functions of the economic system illustrated by Knight's "wheel of wealth," somewhat modified. In these editions Samuelson also makes use of the wheel in order to provide a detailed illustration of the simultaneous, mutual determination of prices in a general-equilibrium framework (1955, p. 597; 1958, p. 615; 1970, p. 607). Thus beginning with the fourth edition of *Economics*, Samuelson presents the wheel-of-wealth diagram at three different points in his textbook, each time in a somewhat different context.

But once again things are not so simple: for neither the *Oxford English Dictionary* nor any of the other standard reference works I have consulted list such a figure of speech. "Wheel of fortune," yes; "wheels of industry," yes; but not "wheel of wealth." Did Knight later delete the words "of speech" from the 1932 version of his discussion because he became aware of this fact? Or was this deletion the result of a slip of the pen or typographical error? Or did a slip occur in the original memeographed version in the very insertion of the words "of speech"? I do not know.

REFERENCES

E. v. Böhm-Bawerk, *The Positive Theory of Capital*, Innsbruck 1888, translated by W. Smart, New York 1891.

E. Cannan, *Wealth*, 2d ed. London 1916.

G. Cassel, *The Theory of Social Economy*, translated by J. McCabe, London 1923; translated, rev. ed. by S. L. Barron, New York 1932.

S. Chapman, *Outlines of Political Economy*, 2d ed., London 1917.

J. R. Commons, *The Distribution of Wealth*, 1893, reprinted, New York 1963.

———, *Institutional Economics*, New York 1934.

M. A. Copeland, *A Study of Money flows in the United States*, New York 1952.

H. J. Davenport, *Outlines of Elementary Economics*, New York 1897.

———, *Value and Distribution*, Chicago 1908, reprinted, New York 1964.

———, *The Economics of Enterprise*, New York 1913.

J. Dorfman, *The Economic Mind in American Civilization*, Vol. 3, New York 1949, Vol. 4, New York 1959.

F. Y. Edgeworth, "Jenkin, Henry Charles Fleeming," in *Palgrave's Dictionary of Political Economy*, Vol. 2, 1899; London 1923, rev. ed., p. 473.

R. T. Ely, *Outlines of Economics*, New York 1893; 6th ed., 1937.

F. Ferrara, "De l'équilibre dans l'échange des productions entre elles," in *Oeuvres Économiques Choisies*, 1864, translated by

G. H. Bousquet and J. Crisafulli, Paris 1938.

I. Fisher, *Introduction to Economic Science,* (Cover reads: *Elements of Economic Science;* actually 1st ed. of next entry), New York 1910.

———, *Elementary Principles of Economics,* New York 1912.

W. Foster, "The Circuit Flow of Money," *Amer. Econ. Rev.*, Sept. 1922, *12*, 460–73.

——— and W. Catchings, *Money,* Boston 1923.

——— and ———, *Profits,* Boston 1925.

C. D. Hurewitz, *Ha-Mammon: Perek ba-Kalkalah ha-Tzibburit (Money: A Chapter in Political Economy),* (Hebrew), Warsaw 1900.

F. Jenkin, *The Graphic Representation of the Laws of Supply and Demand, and other Essays on Political Economy,* London 1887, reprinted London 1931.

E. Kleiman, "An Early Modern Hebrew Textbook of Economics," in *History of Political Economy,* 5, 1973 (forthcoming).

F. H. Knight, [Review] "Cassel's Theoretische Socialökonomie,' *Quart. J. Econ., 36,* 1921–22, 145–53.

———, "*Economics,* 1920's" (unpublished material in Boxes 8 and 9 of Knight's papers that have been catalogued by Glen Gilchrist under George Stigler's direction). In part, early versions of 1932, apparently written during the period 1922–26 (see fn. 1 above).

———, book review of *Profits* by W. T. Foster and W. Catchings, *Polit. Sci. Quart.*, Sept. 1926, *41*, 468–71.

———, "Social Economic Organization," "The Price System and the Economic Process," "Demand and Supply and Price," "Distribution: The Pricing of Productive Services Individually," in *Second Year Course in the Social Sciences (Social Sciences II): Syllabus and Selected Readings,* prelim. ed., Chicago 1932.

———, *The Economic Organization,* Chicago 1933, multilith; reprinted New York 1951. (Reproduces the material of the preceding entry.)

———, *The Ethics of Competition and Other Essays,* New York 1935.

J. J. O. Lahn (pseud. of N. A. L. J. Johannsen), *Der Kreislauf des Geldes und Mechanismus des Social-lebens,* Berlin 1903.

———; N. A. L. J. Johannsen, *A Neglected Point in Connection with Crises,* New York 1908, reprinted, New York 1971.

L. S. Lyon, *Education for Business,* 2d ed., Chicago 1923.

A. W. Marget, *The Theory of Prices,* Vol. 1, 1938; Vol. 2, 1942, reprinted New York 1966.

A. Marshall, *Principles of Economics,* 8th ed., London 1920.

K. Marx, *Capital,* 3 vols., translated from 3d German ed. by S. Moore and E. Aveling, Chicago 1906.

C. Menger, *Principles of Economics,* 1871, translated by J. Dingwell with an intro. by F. H. Knight, Glencoe, Ill. 1950.

W. C. Mitchell, *Business Cycles,* Berkeley 1913.

———, *Business Cycles: The Problem and Its Setting,* New York 1927.

S. Newcomb, *Principles of Political Economy,* New York 1886.

J. S. Nicholson, *Principles of Political Economy,* London Vol. 1, 1893; Vol. 2, 1897; Vol. 3, 1901.

R. H. I. Palgrave, *Palgrave's Dictionary of Political Economy,* New ed., Henry Higgs, ed., London 1925.

D. Patinkin, "Frank Knight as Teacher," *Amer. Econ. Rev.*, Dec. 1973, *63*, 787–810.

E. v. Philippovich, *Grundriss der Politischen Ökonomie,* Freiburg 1897–99.

F. Quesnay, "Tableau Economique, 1758–59," in M. Kuczynski, and R. L. Meek, eds., *Quesnay's Tableau Economique,* London 1972.

M. C. Rorty, *Some Current Problems in Economics,* Chicago 1922.

W. Roscher, *Principles of Political Economy,* Stuttgart 1854; (Eng. trans. Chicago 1878. 13th Ger. ed.).

P. A. Samuelson, *Economics: An Introductory Analysis,* New York 1948, 1951 (2d ed.), 1955 (3rd ed.), 1958 (4th ed.), 1970 (8th ed.).

J. A. Schumpeter, *Economic Doctrine and Method,* 1912, translated by R. Aris, London 1954.

———, "Das Socialprodukt und die Rechpfennige," *Archiv für Socialwissenschaft und Socialpolitik,* 44, 1917/18. Translated under the title "Money and the Social Product" in Int. Econ. Assn., *Int. Econ. Papers* No. 6, London 1956.

E. R. A. Seligman, *Principles of Economics,* London 1905.

P. M. Sweezy, *The Theory of Capitalist Development,* New York 1942.

F. W. Taussig, *Principles of Economics,* New York 1911, rev. ed., 1915.

F. M. Taylor, *Principles of Economics,* 9th ed., New York 1925.

R. D. Theocharis, *Early Developments in Mathematical Economics,* London 1961.

J. Viner, *Studies in the Theory of International Trade,* New York 1937.

A. Wagner, *Grundlegung der Politischen Oekonomie,* 3 vols., Leipzig: Winter, 1892–94.

———, *Theoretische Socialökonomik,* 2 vols., Leipzig: Winter, 1907–09.

———, *Soziälökonomische Theorie des Geldes und Geldwesens,* Leipzig: Winter, 1909. (Vol. 2, part 2 of preceding entry.)

F. A. Walker, *Political Economy,* 3d ed., New York 1887.

K. Wicksell, *Föreläsningar I Nationalekonomi,* Vol. 1, Lund 1901; reprinted as *Lectures on Political Economy,* Vol. 1, English translation by E. Classen, London 1934.

P. H. Wicksteed, *The Common Sense of Political Economy,* London 1910; London rev. ed. 1933.

F. von Wieser, *Natural Value,* 1889, translated by C. A. Malloch, reprinted, New York 1956.

Board of Governors of the Federal Reserve System, *Flow of Funds in the United States: 1939–1953,* Washington 1955.

IN MEMORIAM

FRANK H. KNIGHT

1885–1972

Frank Hyneman Knight was born in White Oak Township of McLean County, Illinois, on November 7, 1885, and died in Chicago on April 15, 1972. He was the oldest of eleven children, of whom nine survived infancy and two (Melvin Moses, of Berkeley and Bruce Winton, of Dartmouth) also became professors of economics.

Knight's early career seems almost aimless in retrospect (and he was fond of saying that he abandoned farming chiefly because his feet hurt). He attended high school in Lexington, Illinois (1903–05), American University in Harriman, Tennessee (1905–07), was a stenographer at the Jamestown Exposition in Norfolk, Virginia (1906–08), attended Milligan College in Tennessee (1908–11, Ph.B. 1911), then University of Tennessee, 1911–13 (B.S., M.A. 1913), where he was an assistant in chemistry, and wrote a master's essay on "Gerhard Hauptmann as an Idealist," and finally Cornell University (1913–16; Ph.D. 1916). He was an excellent student (even in deportment, in which he received 100 each year at Milligan!), but not free of skepticism from his earliest years. In his autobiography (*Pioneer's Progress*, 1952), Alvin Johnson reminisced about Cornell days:

> Knight came up from Tennessee, where border-state diet had endowed him with dyspepsia and a graven expression of pessimism. He was majoring in philosophy under Creighton and Thilly and doing a minor with me. I found him the keenest student of theory I had ever had. When I was elaborating some theoretical subtlety he would turn on me the gray light of his skeptical eyes, seeming to say, "I don't believe a word of that." Sometimes he would attack my position, very competently. I liked that. It was stimulating to wrestle over points of theory.
>
> One day in the spring Knight came to me with an unusually sad face. "Doctor," he said, "I thought all along I was doing pretty well in my philosophy major. But Creighton called me in and gave me the most frightful hauling over the coals. Said I was totally unfit to study or teach philosophy; indeed, that I ought to drop the idea of teaching anything, for with my attitude I'd do more harm than good. I went to Thilly for comfort, and he gave it to me still worse."
>
> I was astonished. Could two such clear-headed philosophers make such a mistake about so promising a student as Knight? I couldn't believe it, I went straight to Creighton.
>
> "Knight?" he said bitterly. "It isn't that he is devoid of ability. But with his ingrained skepticism he repudiates all the values of philosophy. As a teacher or writer he will be not just the blind leading the blind into pitfalls. He will destroy the true philosophic spirit wherever he touches it."
>
> "I see," I said. "Knight has been miscast. He should not have majored in philosophy, where his work and attitude are so unsatisfactory, but in economics, where he does excellent work. Suppose I take him over?"
>
> "Do as you please," said Creighton acidly. "We in philosophy will have nothing to do with him."

The years 1913–16 marked the rapid mastery of economic theory. In the former year, Knight had essentially no knowledge of the subject; in the latter year he had completed *Risk, Uncertainty, and Profit* (with the title, *Cost, Value, and Profit*), one of the most influential doctoral dissertations in the history of American economics. With revisions, it won second prize in the then-annual competition of economics, the Hart-Schaffner-Marx prize for 1917 (the first prize was won by E. E. Lincoln, *The Results of Municipal*

Electric Lighting in Massachusetts), and it was published in 1921. The depth of the knowledge of economic theory that Knight acquired in these few years is simply incredible.

After a year as instructor at Cornell, Knight moved to the same post at Chicago (1917–19), and then became an assistant professor and (after 1922) professor at Iowa. He resigned from Iowa in 1928, and spent the rest of his life at the University of Chicago. He became emeritus in 1955 but continued to teach until 1958.

He married Minerva Shelburne in 1911, and they had three daughters and one son. They were divorced in 1928. In 1929, Knight married Ethel Verry and they had two sons.

Risk, Uncertainty, and Profit was not only Knight's first systematic treatise, but also his last. This remarkable work which was reprinted many times in the London School series, not only influenced the thinking on risk and profit of a generation of economists but also served as one of the major treatises in graduate courses on economic theory. The following decade was filled with articles of the highest distinction—most of the best of them included in *The Ethics of Competition* (1936). The essay on the nature of cost, directed to Pigou's famous error on increasing cost industries, is perhaps the most lasting. In the 1930's, his interests in economics shifted to capital theory, and his famous debate with the Austrians (Machlup and Hayek, and the temporary Austrian, Kaldor) covered the decade.

Beginning in the 1930's, however, his main interests began to shift from economic theory to social philosophy: the nature of public discussion and decision making; the content (or absence thereof) of religious ethical systems; the contradictions in democratic planning. A profound distrust of democratic politics (which reached its climax in the dark days of 1932, when he delivered his lectures on "The Case for Communism") was joined to an even more profound fear and dislike of all authoritarian systems: Knight was deeply pessimistic on the rationality of all political processes.

Frank Knight was one of the last prominent economists to avoid all significant connections with the political and business worlds. Even the presidency of the American Economic Association, nomination to which he declined for 1937 and 1938, was complimented as burdensome and criticized as diversionary, although he was persuaded to accept the office in 1950. He received the Francis A. Walker Medal in 1957. For many years he served as coeditor with Jacob Viner of the *Journal of Political Economy*. The impressions made by his highly idiosyncratic teaching have been discussed by two former students (Stigler, *JPE*, May–June 1973; Patinkin, this *Review*). For a full generation he was the dominant intellectual influence upon students at the University of Chicago, and in general his influence was most profound upon the most able students. A reasonably complete bibliography through 1936 is contained in *The Ethics of Competition*. Two other extensive collections of his essays are *Freedom and Reform* (1947) and *History and Method of Economics* (1956). He translated Max Weber's *General Economic History* (1927), and wrote with T. W. Merriam, *The Economic Order and Religion* (1945).

Combatting Role Prejudice and Sex Discrimination

Findings of the American Economic Association Committee on the Status of Women in the Economics Profession*

I. Role Prejudice as an Economic Problem[1]

It is well recognized among economists that arbitrary discrimination in the labor market creates a general economic loss. Where certain people are excluded from jobs for reasons which have nothing to do with the performance of the jobs themselves, the loss to consumers plus the loss to those who are excluded is greater than any gain to those who are favored. There is now copious literature on this subject.

It is less well recognized that discrimination among existing members of the labor force is only a special case of a much larger process of role learning and role acceptance, which begins almost from the moment of birth. It is not merely that differences in skills are learned, as in Adam Smith's famous passage about the porter and the philosopher,[2] but images of possible roles on the part of both the role occupants and the role demanders are likewise learned in the long process of socialization. The existing division of labor at any one time, therefore, may reflect "role prejudice"—that is, a learning process by which certain irrelevant biological or genetic characteristics of individuals are associated with certain roles. Some genetic distinctions are significant for role performance, but many are not and yet are widely believed to be significant. These beliefs are products of a false social learning process, which can be a source of economic loss even larger than false discrimination within an existing labor force. Discrimination in an existing labor force indeed emerges as one consequence of the much larger process of the development of role prejudice.

Role prejudice tends to develop when there are genetic differences in the human population which are visible but are not necessarily significant for role performance. We see this clearly by considering those genetic differences which are undetectable except by refined methods and therefore are socially invisible. Some of the blood types are perhaps an example. In ordinary life we are quite unaware of a person's blood type. Consequently, it can be safely assumed that there is no role prejudice against blood types of different kinds, unless these happen to be associated with other characteristics which are socially visible. For the major types, this association is fairly minimal. Consequently, we would be very surprised to find a distribution of blood types in any occupational group or income group of any size which is markedly different from the distribution of blood types in the population as a whole.

For any particular distribution of genetic characteristics, it is clear that we can postulate "no-prejudice proportions" in all the different occupations and structures of society. In the case of the socially invisible and irrelevant characteristics mentioned above, the no-prejudice proportions, subject to random variation, in any group of society would be expected to be the same as the proportion of the characteristic in the society as a whole. If there were any case—and I must confess I have not been able to think of one—of a genetic characteristic which is relevant to role performance but which is socially in-

* This report is published pursuant to a resolution adopted by the members at the Annual Meeting of the American Economic Association, December 28, 1971. In providing for establishment of the Committee on the Status of Women in the Economics Profession, the resolution specified that "Its general findings, conclusions, and recommendations shall be published by the American Economic Association upon the Committee's request" (*Amer. Econ. Rev. Proc.*, May 1972, p. 474). This report has not been considered or approved by either the Executive Committee or the membership.

[1] Part I was drafted for the Committee by Kenneth E. Boulding, professor of economics, University of Colorado.

[2] "The difference between the most dissimilar characters, between a philosopher and a common street porter, for example, seems to arise not so much from nature, as from habit, custom, and education" (p. 15) Adam Smith.

visible and not in the information system, then the existing structure of proportions of the characteristic in different occupations might not be the same as the proportions in society as a whole, and those occupations which this invisible characteristic favored would have a larger proportion than in the society as a whole, but this would still constitute a no-prejudice distribution. Paradoxically enough, without knowledge there can be no prejudice; without some knowledge, there can be no false knowledge.

Role prejudice may easily arise where there are highly visible and observable genetic characteristics, such as sex, race, height, size, color, and so on. Historically these have frequently been perceived as having some relevance to performance in different roles. But this relevance tends to be exaggerated, because a rational tendency to economize information can easily turn into an irrational process of stereotyping and false knowledge.

The genetic division of the human race into men and women qualifies in most societies as the major form both of discrimination in the existing labor force and of role prejudice. This is not surprising, as the genetic difference between the sexes is far greater than it is among the races in terms of the structure of information of the genetic code. There is one role indeed—that of bearing children—in which there is 100 percent genetic specialization between the sexes, in the sense that no men will be found in the child-bearing population. The very social visibility of the distinction between the sexes, however, makes it a prime candidate for role prejudice, which creates a huge reservoir of false knowledge, false education, and false learning processes. All are costly to society in exactly the same sense that an unknown and unutilized natural resource is costly in the sense of opportunity foregone. We can, oddly enough, think of role prejudice as a large natural resource which can be mined by the development of true knowledge and better learning processes to the benefit of the whole society.

Where the sex distinction is completely irrelevant to role performance, role prejudice can be said to exist wherever there are substantial deviations from the 50 percent rule —that is, 50 percent of each sex in the occupation—this being assumed to be the proportion of the sexes in the total society. The difficult case, of course, is where there is some genetic foundation for partial role differentiation. It is the cases between the 50 percent and 100 percent that are difficult, and in these cases there is often no information system—at the moment at any rate—which can inform us as to the exact proportion of the sexes in different occupations which constitutes zero role prejudice. Fortunately, however, we do not have to know where the exact position of the ideal proportion is if the existing proportion clearly represents a gross disproportion. In order to set up a social policy we do not have to know exactly where the ideal lies. We merely have to identify those cases in which the deviation from the ideal is so clear and striking that we can find wide agreement as to the desirable direction of change. At the bottom of the mountain, one knows which way is up, even though one may not know exactly where the top of the mountain lies. Maximizing is an illusion, but betterment is not.

It is the contention of the American Economic Association Committee on the Status of Women in the Economics Profession (the Committee) that there is gross disproportionality in the proportion of women to men in the economics profession, and especially in university teaching, as revealed in the data which have been collected. Fortunately, we do not have to argue whether the genetic differences between women and men would result in a no-prejudice proportion of women in economics of 40 percent or 50 percent or 60 percent when the actual proportion is less than 10 percent. This gross disproportionality, of course, arises from sources in the long history of the human race, most of which are outside the economics profession. This, however, does not excuse the economics profession from setting in motion processes which will raise its proportion of women. We are a sector of the knowledge industry and a group of people concerned with eliminating sources of unnecessary economic loss. Only if we are successful beyond our wildest expectations, considering the vastness of the mine of ig-

norance and false learning, will it be necessary to raise the question as to whether we have gone too far.

The critical question, therefore, is that of the "betterment production function," as it might be called. The social indicator of betterment is in this case clearly an increase in the proportion of women in the profession, especially in those branches where the proportion is least. If the rate of change of this proportion is regarded as the product, the question is: What are the inputs which produce this output, and particularly, what are those inputs that can be most easily expanded and that have the highest marginal productivity? We might also look at those inputs which have a high negative marginal productivity and which can be contracted. Four broad classes of inputs may be named: Information, Persuasion, Rewards, and Punishments.

1. *Information*

Information properly organized changes the images of the world of people who receive it and consequently changes their behavior. There are two aspects of the information process— theoretical structures and information about fact. Better theoretical structures are an essential ingredient in the advancement of knowledge, and in this particular case it is important that the theory of role prejudice, as outlined above, should be widely recognized as a legitimate extension of economic and social theory. One of the greatest enemies of the fact of prejudice is the spread of the idea of prejudice. Once it is recognized as a theoretical construct, the facts of perverse social learning which previously have been invisible become glaringly visible. The improvement and elaboration in detail of the theory of role prejudice is still an important task for the future.

Coupled with a wider appreciation of the theory, the development of statistical information, case studies, and a continuing series of social indicators can serve as a constant criticism of what is false in existing theories of the world and a reinforcement of what is true. It is tempting to think of every advance in knowledge as the result of a combination of new theoretical insights with new information collection and processing procedures. In social systems, however, knowledge does not merely reflect the world—it changes it. And in this case particularly, the more we know about role prejudice, the less there is likely to be. Getting lost is one way to improve one's knowledge of the terrain, but studying a map is a better way, and diminishes the amount of getting lost.

2. *Persuasion*

The mere existence of knowledge and information is not always enough to insure its spread. One of the major sources of ignorance is inattention, and to insure the spread of new knowledge it is often necessary to call it to the attention of people in ways which are dramatic and persuasive. The fact that persuasion can be abused to propagate false knowledge does not excuse us from its proper use in the propagation of truth. Persuasion plays a particularly important role in the change of values, and without getting into the ancient controversy about the truth of values one has to recognize that change in our image of the world inevitably changes our values about it. As scientists, we have to guard against the information filter which our values create. We cannot pretend, however, that information is value free. Change in knowledge and change in values are joint products of a single process.

A problem of particular difficulty is that of irrelevant and unconscious sources and attitudes, even in the academic and intellectual community, which is supposed to be on its guard against irrationality. The relations between the sexes are often characterized by high levels of emotion, and hence are particularly subject to the distortions which arise from unconscious or irrelevant sources. It is one of the functions of ethical persuasion to raise our awareness and to set up defenses against attitudes and prejudices which arise from this source.

3. *Rewards*

It is widely recognized in the learning process that positive and negative payoffs are of great importance. Every economist is familiar with the proposition that any human activity which is rewarded above some

neutral level will expand and those rewarded below this level will contract. The structure of knowledge and values in the human mind, likewise, grows towards its own more positive payoffs and away from the less positive and negative payoffs. It is by constructing a reward structure that society is most likely to achieve change. These rewards may be either internal or external. Our own applause, as Paul Samuelson suggested, may indeed be the greatest reward of all. No economist, however, can deny the significance of external reward, whether the applause of others or those more tangible things, such as subsidies, tax remissions, and price changes. Curiously enough, although psychologists seem to agree that rewards are more successful than punishments in the learning process, it seems harder to legitimate a reward structure. A suggestion, for instance, that dues paid by members of the American Economic Association should be scaled down if the organization or department to which they belong had a high proportion of women might be hard to legitimate, even though it might be effective. Nevertheless, the search for reward structures should go on.

4. *Sanctions and Punishments*

Punishments always seem to have been easier to legitimate than rewards, perhaps because of a belief that they are cheaper, in spite of much evidence that they do not contribute as much to learning. Nevertheless, there is one extremely good reason for the use of sanctions. In the case of public goods and public bads it is well recognized by economists that legitimated threat of some kind is essential if we are to avoid some "tragedy of the commons." Private reward structures, such as the market and the price-profit mechanism, are particularly well adapted to the provision of private goods, but are poorly adapted to public goods, where without some sanctions in the collection of taxes public goods would simply not be provided in adequate amount. Private reward systems are also poorly adapted to dealing with public bads such as pollution, of which role prejudice is a prime example, in the sense that everybody is somewhat diminished by it. The "freeloading principle" applies here, for if there are no sanctions it will pay the individual not to participate in a collective sacrifice required for the increase of a public good or the diminution of a public bad. To rely on purely private rewards and exhortations for the diminution of role prejudice is rather like trying to finance a government with purely voluntary contributions.

To say that sanctions are necessary does not mean to deny that there are acute problems both in the quantity and the quality of sanctions, and that bad sanctions may be worse than no sanctions. This, however, does not excuse us from the search for good sanctions, which must not be larger than necessary, must be credible, legitimate, precise, equitable, and so on. Sanctions should fall on violators and not on the innocent; they should be appropriate—that is, any punishment should fit the crime—and it should be recognized that, like all deterrents, punishments are most successful when they are not carried out. Economists will recognize, however, that no system of sanctions can be perfect and that there has to be some kind of tradeoff between imperfection in the system of sanctions which will inevitably lead to some failures of both justice and of deterrence, and the failure of the provision of public goods or the elimination of public bads which too weak a system of sanctions might involve.

The power of the American Economic Association to invoke sanctions is small indeed, and it could be argued that, apart from publicity—which under some circumstances is itself a sanction—a professional association should not legitimately engage even in linguistic sanctions like censure. Nevertheless, it is only realistic to be aware of the fact that sanctions of government can be invoked by groups in the society in the furtherance of what they believe to be general betterment. Even a profession is not exempt from participating in the society of which it is a part. It is important that vigilance be maintained against both the abuses of sanctions and their undue neglect, even though the primary

emphasis of the intellectual community is properly on information, persuasion, and rewards.

The Committee is aware that it has a special task which is part of a much larger social process. Role prejudice in particular is by no means confined to women. Men also suffer from role prejudice, and it accounts for other kinds of discrimination than one based solely on sex. Nevertheless, as economists we believe in the division of labor and we believe that the pursuit of our particular task is to complement rather than compete with the labors of others in the long process of social betterment. The guidelines that we are proposing, while they are particularly applicable to our special case, are all simply examples of the fundamental general principle—that where there is clearly not enough, the simplest and most necessary guideline is "more." In what follows we spell out in detail this essential principle.

II. Guidelines to Obviate Role Prejudice and Sex Discrimination[3]

The following guidelines are written for members of the American Economic Association who, as faculty members in departments of economics in various colleges and universities, as economists in business, industry, or government, wish to remove, mitigate, or prevent discrimination against women economists. These suggestions have been prepared by the Committee to illustrate steps that can be taken by those of good will who seek to comply with the spirit as well as the letter of laws proscribing employment discrimination. Over time, other suggestions may be appropriate; and the guidelines may need amendment or modification to reflect changes in the conditions of employment faced by women economists.

Not all institutions will wish to implement all the recommendations given here. Nevertheless, thoughtful consideration of whether a particular suggestion has merit for a given institution may reveal needs heretofore unrecognized and may be the basis for setting in motion policies that will eliminate sources of discrimination and unnecessary economic loss.

Because many members of the American Economic Association will wish to review again the resolutions relating to elimination of sex discrimination among economists that were passed by the Association at the annual meeting in December 1971 at New Orleans, the resolutions are repeated here.[4] They provide the framework for the Committee's guidelines for positive policies and remedial action to obviate role prejudice directed against women economists.

> I. RESOLVED that the American Economic Association declares that economics is not exclusively a man's field. The Association herewith adopts a positive program to eliminate sex discrimination among economists, whether employed in universities and colleges, industry, finance, publishing, or other endeavors. The President shall make known, by all available means, the Association's adoption of the following principles:
>
> a) To redress the present low representation of women in the economics profession, the Association shall actively encourage the study of economics by women at all levels of education.

Low representation of women in the economics profession occurs in part because women students do not have appropriate role models at the university nor in visible positions in other places of employment. Women also lack accurate information and encouragement throughout the entire educational process. The role models needed are professional women economists with whom young women can identify and from whom young women can obtain information and counseling. It is also important that young men in the classroom recognize, as a matter of course, that professional women play a vital

[3] Part II was written by Barbara B. Reagan, professor of economics, Southern Methodist University, on behalf of the Committee. These guidelines reflect the input and discussion of all Committee members as well as public discussion in Toronto at the December 1972 meetings of the American Economic Association.

[4] The resolutions are contained in the Appendix to the minutes of the annual meeting, pp. 473–74.

role in economics as well as other fields, and that professional competence among adults should not be sex-related. Throughout their entire training, children, both boys and girls, should have opportunities to appreciate the economic contribution of women to our society, and the economic circumstances of discrimination against women. By the time young women have completed high school, they should know that job planning is essential for everyone and should understand that economics is a perfectly appropriate field of study for women.

There are numerous things that economists can do to encourage the changes needed to overcome previous false social learning.

(1) *Educational Institutions.* The value of women faculty role models at the college or university level cannot be stressed too much. A substantial amount of talent that would otherwise flow into this profession is lost because, without the presence of women faculty, undergraduate women erroneously assume that economics is a profession for men only. An exclusively male department also lacks the abilities to correct these impressions, not least because men economists themselves overlook the difficulties subsumed under the cliché "economics is a man's field." A women economist can provide encouragement and proof that it is possible not only to survive but to accomplish, even as a member of a tiny minority group. The positive role that women faculty can play in enhancing the education of both women and men should be considered in making appointment decisions.[5]

Departments should also use other methods of making women more visible—inviting guest lecturers (for example, women from government, industry, management, finance, public action programs, research), hiring women visiting professors, featuring women in career opportunity conferences, using women as special lecturers, recommending women experts when students want someone with whom to consult, and including women economists in informal gatherings of economists. Again, these are ways of doing away with role prejudice.

Departments should facilitate communication between senior women economics majors and freshmen women so that the latter may be appropriately informed and encouraged to study economics. Whenever possible, communication between women graduate economics students and women undergraduate majors should be similarly facilitated.

(2) *Government.* Part of the lack of role models for potential women economists has arisen because the government at all levels has put little effort into appointing women economists to highly visible positions. Economists should work to increase the number of women economists in high ranking positions in the federal and state governments to serve as examples and encouragement to women facing career decisions.

The federal government, as part of the Federal Women's Program, should require that women be given equal opportunity in government-financed graduate student programs and should encourage women in such programs to study economics.

(3) *Private Industry.* Private industry should search for women economists in order to place them in both research and administrative positions. The reciprocal employment relationships between industries and universities (and/or consulting arrangements) should be encouraged for women economists as it has been for men.

Training of mid-career women economists already employed in the private sector should be encouraged. Both full- and part-time training and educational opportunities should be made available to women economists, as to men, to enable them to acquire new skills and improve their status

Resolution Ib) No academic institution or department shall discriminate against

[5] Elizabeth Tidball has established a correlation between the number of women faculty at a school and the number of women students who subsequently attain recognition for their achievements, using a random selection of listings in *Who's Who of American Women* for a three-decade period. The statistically significant positive correlation coefficient was $+0.953$ with $P<0.005$. By contrast, "the number of men faculty neither enhanced nor detracted from the output of women achievers."

women in admission to studies in economics or in the form or amount of financial aid. Every economics department shall actively encourage qualified women graduate students without regard to age, marital or family status, and shall actively promote their subsequent employment.

(1) *Admissions.* The admissions process should be checked by economics departments to ensure that there is no discrimination on the basis of sex, age, marital, or family status. In efforts to eliminate or preclude such discrimination, whenever possible, women faculty or graduate students should participate in the admissions process. When prospective candidates for admission are interviewed, women applicants should not be asked questions about marital or family status (actual or prospective) that are not asked of men.

(2) *Recruitment.* Because the applicant pool is often low, departments of economics should actively recruit women into their programs at both the graduate and undergraduate levels. Women graduates of the department form a good supplement to such recruiting efforts. To give evidence of career opportunities, departments might keep a file of letters from former economics majors on their experiences with further academic training and/or the labor market. Whenever possible, women graduates from the department should be invited back to give lectures. When recruiting graduate students, particular attention should be paid to institutions from which large numbers of qualified women may be recruited; such places may include state and city colleges and women's colleges. Letters to colleagues teaching undergraduates should publicize the department's interest in women applicants. When considering applications, women with undergraduate preparations that are closely related to economics should also be considered.

(3) *Financial Aid.* When faculty members or department chairmen are making decisions with respect to the size of grants or fellowship aid for various graduate students, marital status should have no relevance. In particular, married women should be treated in the same way as married men. Under no circumstances should it be assumed that a married woman's needs are less than a man's without due consideration of the financial facts involved.[6]

Information on sources of fellowships and grants should be distributed so that all students, both men and women, receive the information on an equal basis. When departments are asked to make nominations for financial aid, qualified women should be included.

Federal scholarship and loan programs should actively seek out women graduate students as potential recipients and should inform all departments of their interest in women applicants.

(4) *Encouragement in the Program.* Those departments that at present have no women faculty members should make special efforts to schedule guest lectures by women economists so that graduate students will be more aware of the active participation of women in the field. Faculty should include publications of women economists on reading lists, particularly in basic courses.

(5) *Placement.*[7] In writing recommendations for prospective employment, econo-

[6] A relevant court case in 1971 involved John and Mary:

"... high school sweethearts who married shortly after graduation. Mary went to work for a bank in a small city in Northern Indiana. John worked as a salesman in the same city. Mary prospered, was promoted, and eventually became a cashier. John did less well. Growing restless, he enrolled in college, using Mary's earnings as the basis of his support. Eventually he graduated with a Master of Science degree in Science Education and entered the teaching profession. In college, and after graduation, reportedly because of college experiences, John claims that he grew intellectually and socially while his wife stagnated. Finding Jane, a fellow teacher, more compatible, he fell in love and pursued an affair with her. Mary sued for divorce and claimed as part of her settlement the present value of the college education of her 41 year old male Caucasian husband, the value of which was computed at $49,653. In the face of this suit John saw Mary in a new light. They were reconciled."

(This information was provided to the Committee by David R. Witmer, Assistant Chancellor, University of Wisconsin-La Crosse.)

[7] Although these comments are written with respect to recommendations for and placement of women graduate students, they apply equally to placement of women economists in subsequent stages of their careers.

mists should not include comments on physical appearance; such comments are not relevant in evaluation of a candidate's qualifications for employment. In particular, care should be taken that comments on physical appearance are not used differentially in evaluating female and male applicants. References to so-called "feminine" or "masculine" traits should be omitted. Marital status, number of children and other private facts of a candidate's life should not be assumed to affect a person's ability to meet professional obligations. Comments regarding political views, particularly the Woman's Liberation Movement, are appropriate only insofar as they are relevant to the applicant's field of research or scholarship.

Care should be taken not to use traditional sex-related stereotypes—downgrading women for conforming to the "passive-accepting" model, or condemning women as unduly "aggressive" for competing to a degree well within the bounds of normal behavior set by their male colleagues for themselves. The latter type of participation may seem unduly aggressive only when measured against a passive, accepting stereotype of the feminine role, but not when measured against professional standards.[8]

Resolution Ic) No employer, academic or otherwise, may discriminate against qualified women economists as candidates for positions, or for promotion or tenure, or for assignment of duties and responsibilities.

Employers should actively recruit women for the economics staff so as to increase diversity—sex as well as race, field or specialty, and school of training. Evaluation of employees' performance and potential for advancement should be based on the same criteria for women as for men. Assumptions about availability for employment as economists should not be made on the grounds of marital status, other employment, home obligations, or the like.[9] Women faculty members should share equally in administrative and committee appointments. Temporary positions or consultantships should not be used for women as a subterfuge to avoid naming them to permanent positions. Titles such as lecturer or adjunct should not be used to deny tenure track positions. Similarly, in government and industry separate job classifications should not be used as a subterfuge masking discrimination against qualified women economists.

All women should have access to management and executive training courses. In-service training opportunities, "acting" administrative positions, and other forms of "grooming" for advancement should be made available for women as well as for men. Informal benefits, as for example, travel reimbursement, conference invitations, and speaking opportunites should be equitably apportioned. Government agencies should submit evidence to the Civil Service Commission that women are being included in these activities. Follow-up evaluations of training courses should include information as to whether women participants in such training courses move ahead in their jobs as rapidly as men.

Members of the American Economic As-

[8] "Women candidates, to win entire approval, have to be both chic and brilliant, and so the woman who is plain-looking, a spinster-scholar type, evokes a negative response for she is not a complete woman, she lacks sociability and will not flatter the egos of male department members. Dossier after dossier divulges irrelevant, negative commentary: '—is a large, broad-boned, somewhat awkward young woman who must be close to six feet in height,' 'her mousiness belies a sharp mind'; '—is . . . tall and proportioned like an Olympic swimmer'; 'she is a steady woman who will never marry.' And within a single paragraph from a male at an elite graduate school; she has a 'comfortably upholstered' person and personality, she performed 'athletically' in a particular course, she would be the 'wheelhorse' of any committee on which she served. And of an older woman, 'if she has any faults, they are those that usually accompany the ambitious woman of her age.' In other words, the rare comment on a male's appearance is simply a footnote, the frequent comment on a female's, a thesis statement." . . . "Our experience of women's dossiers confirms that a letter writer, often unselfconsciously, diminishes a female candidate's intellectual power by stereotyping her as too feminine, too pretty. These comments, a composite from a type candidate's dossier, make a damning package: sweet, but not saccharine, quiet, unaggressive, shy, but very pretty, a decoration to the classroom." [M.L.A. Newsletter, p. 3]

[9] ". . . the employment form of a major oil company included the question, 'can this employee relocate?' On examining the completed forms of 300 women employees, the company officials found the appropriate box marked 'no' in every case. On probing further, they discovered that no one had actually asked the women whether they would relocate; it had simply been assumed that they would not." [Boyle; p. 94]

sociation should review written and unwritten membership criteria of local or regional economics associations to ensure that women economists enjoy equal access and equal opportunity for participation. Members of the Association can be effective in protesting the exclusion of women from membership in any local economic association or any *other professional organization* to which they belong that is not willing to invite women economists to be members. Membership in such organizations is an important part of the informal professional network, and denying women access to any part of the network is sex discrimination.[10]

Positive efforts to regularly include women economists in the informal professional network should be conscientiously encouraged by all. The inclusion of women in this network should be considered a matter of routine operating procedures.[11]

Resolution Id) Salaries, fringe benefits, and facilities and resources for research for women shall be the same as those for men in the same position and rank.

All employers should regularly compare the salary, training, fringe benefits, and other terms and conditions of employment for women economists with those for male economists with comparable education and experience. Where women have not received treatment similar to that given men, corrective action should be taken immediately. Remedies may include giving the woman assistance with "upgrading" herself through leaves or other devices, changing rank or position, increasing base pay and granting back pay. (That back pay should be accompanied with interest may seem an ideal to some professionals, but scarcely to economists.) Rectifying a previous inequity can in no way be construed as reverse discrimination. Indeed, the aim of the action normally has to be to change the initial distribution of emoluments in favor of the party subjected to discrimination. By definition, at least the relative shares of all other groups must decline. There is no other way to correct unfair shares. What happens to absolute shares obviously depends on what's happening to the size of the pie.

After corrective action is taken, care must be exercised that equal treatment is maintained in the future. It is very easy to recognize that an equity raise has been given a women, and then during the following years to slip into old patterns of thought so that merit raises gravitate to men in the department, thus reestablishing old patterns.

(1) *Salaries.* Employers should review and evaluate salaries paid to women even though no formal complaint or charge has been filed. Conferences with the women economists involved may prove helpful in diagnosing situations in which women feel that their salary is inappropriate. Lower geographic mobility among women should not be assumed or used as a rationalization for lower salary. Employers should not base the pay of a new employee solely on the basis of previous pay. The federal government should not continue to base GS ratings for entering academic economists primarily on the basis of total salaries before entering government service. Rather, the GS rating should be based upon objective criteria relevant to the job description. Otherwise, discrimination against women economists entering government service will simply compound past discrimination resulting from denial of equal access to outside consulting and related activities.

(2) *Fringe Benefits.* Employers should review all fringe benefits for possible sex discrimination and ways to facilitate employment of women. Examples include health and life insurance, pensioning, tuition grants for

[10] In 1921 a woman graduate student at Stanford joined her professional association. When she married, she kept her membership; her husband was not then a member. Some years after that it seemed sensible for him to belong to the association and they made it a "family" membership, which they continued through 1972. In December they resigned their memberships. After some weeks the husband received a letter from the Secretary's office expressing regret for his resignation after having been a member for so many years and offering him a membership without publications. No such letter was sent to his wife although she had been a member for over half the century. (This information was presented to the Committee by Lucy Winsor Killough, a retired member of the Wellesley College faculty.)

[11] For many years, the Yale Economics Department conducted official business at a male-only club.

children, mortgage or other loans at low interest rates, subsidized housing, child care facilities on campus and at conferences, recreational facilities, and various sorts of leave. Women should not be discriminated against because of an actual or potential maternity leave. Men and women economists should be offered the same facilities and resources for research, including assistants, offices, and equipment.

Resolution Ie) There shall be no distinction between men and women as to what is considered to constitute full-time employment.

Both part-time and full-time positions should be described completely with respect to the obligations of the persons appointed. For example, many academic appointments consist of a combination of duties: teaching, consulting, advising students, departmental governance, recruiting, supervising personnel, conducting research, representing the department, and so on. If one or more of such activities is optional, performance in the optional activity should not carry any weight in determining promotion or compensation. Those duties that are a necessary part of the job should be clearly stated and a minimum time should be specified.

Some women may wish part-time appointments, particularly for the few years while their children are small. Others may wish full-time appointments. Preference of individuals should be considered. Institutions should not base their hiring policy on the myth that most working women are interested in part-time work primarily to earn pin-money. On the other hand, institutions should increase the part-time options for those, men or women, who are interested. (See Resolution III.)

Care must be taken that "part-time" is not another job title that is used as a convenient way to deny job opportunities for women who wish to work, or in fact do work full time.

Resolution If) Employment may not be refused to any qualified economist on the grounds of family relationship to another employee.

Decisions regarding appointment, promotion, salary, or tenure should be unaffected by the presence of a close relative in the same or related firm, department, or government agency. However, employees should neither initiate nor participate in institutional decisions involving a direct benefit (initial appointment, retention, promotion, salary, leave of absence, etc.) to members of their immediate families.

Much of the past discrimination against women has been related to their marital status as well as their sex.[12] Economists employed by a college or university, a firm, or a government agency with practices in conflict with those suggested here should campaign actively to have the rules or practices changed.

Without compromising standards in any way, special efforts should be made to facilitate the employment of couples who apply to institutions as teams. Such employment should be in keeping with the individual talents of the team members so that neither part will be specially burdened.

III. RESOLVED that the American Economic Association shall encourage flexibility in providing opportunities for education and employment on a part-time basis both for men and women.

Departments of economics should review all educational requirements that include a time specification. It is expected that a conscientious reexamination of existing procedures will lead to considerably more flexible

[12] "The invisibility of the female job candidate emerges nowhere more unselfconsciously than in the dossiers of young, married graduate students. The dossiers of undeniably brilliant, serious, promising women, filled with letters from eminent advisers, very frequently compare the woman with her husband, stereotype her as the teacher not the scholar-thinker, and so the less valuable member of the couple. Indeed, one often wonders for whom the letter was written. One woman's letter contains this sentence: 'Mrs. —'s husband — is an excellent scholar, rather more disciplined and professional . . . . Mrs. — complements him with her more imaginative and enthusiastic . . . literary rather than historical perspective.' . . . Or another begins: '— and her very able husband' and continues 'like her husband, she . . .' " Such letters affect the reader by reinforcing his notion that a woman is an appendage, her career second to her husband's; her talents shine in the light of his. [M.L.A. Newsletter, p. 3]

arrangements at many academic institutions.

The question of part-time employment is a matter of department or institutional policy, but the question of whether the policy applies equally to men and women is not. It is imperative that in considering whether to allow (or encourage) part-time appointments, only the impact on work activities be considered, and not the reasons why people wish part-time appointments. If policies on part-time appointments enable an employer to secure an expert who will teach one course or do one project although engaged elsewhere in other paid employment, then part-time employment should also be available to women—or men—engaged in work ordinarily considered a nonmarket transaction. The person who carries a portion of a conventional full-time load should have assigned duties and base compensation (full-time equivalent) comparable to those of full-time appointments, and duties, salaries and fringe benefits should be prorated. Academic departments should particularly look at such fringe benefits as research assistants and travel grants which tend to increase professional competence.

The part-time person who teaches one course or does one project may be hired on a consulting basis for a flat fee. In such cases no fringe benefits need be available if the fee is adjusted accordingly. Economists at academic institutions can encourage and assist nonacademic organizations to increase part-time opportunities even if department policy has ruled against this for faculty positions. Credentials, skills, and professional creativity required of part-time personnel should be at least equivalent to the levels required of full-time employees so as to justify the equivalent base pay. Part-time appointments should not be a way to get cheap, low quality service. Nor should they be used as means of psuedo-compliance with Affirmative Action Plans.

The Committee is willing to discuss any of the above guidelines in the context of particular situations or additional suggestions of ways to improve current employment conditions. Members of the American Economic Association should report to the Committee on the Status of Women in the Economics Profession any cases that appear to involve sex discrimination directed against women economists. Members are reminded that the Wage and Hour Division of the U.S. Department of Labor as a matter of course receives complaints of wage or salary discrimination on a confidential basis, with the possibilities of back pay being awarded. Furthermore, the Equal Employment Opportunity Commission acts on complaints of unequal advancement opportunities for women.

III. Summary of Action Responsibilities

To comply with the resolutions adopted by the American Economic Association on December 28, 1971, a summary of action responsibilities, italicized below, has been prepared. Their central thrust should command virtually unanimous support: to eliminate extraneous considerations such as an individual's sex from any judgment of ability, performance, or potential.

I. RESOLVED that the American Economic Association declares that economics is not exclusively a man's field. The Association herewith adopts a positive program to eliminate sex discrimination among economists, whether employed in universities and colleges, industry, finance, publishing, or other endeavors. The President shall make known, by all available means, the Association's adoption of the following principles:

a) To redress the present low representation of women in the economics profession, the Association shall actively encourage the study of economics by women at all levels of education.

ACTION RESPONSIBILITY: *It is our obligation to affirm, in precept and action, that there are no genetic distinctions between the sexes which would preclude the study of economics by women on a proportionately equal basis with men. Since the role prejudice on the part of role occupants and role demanders is an acquired (i.e., learned) characteristic, educational institutions, government, and private industry should help dispel the prevailing myth that the current division of labor in the*

economics profession reflects relevant biological or genetic differences between the sexes.

b) No academic institution or department shall discriminate against women in admission to studies in economics or in the form or amount of financial aid. Every economics department shall actively encourage qualified women graduate students without regard to age, marital or family status, and shall actively promote their subsequent employment.

ACTION RESPONSIBILITY: *With respect to admissions, financial aid and placement, the primary concern should be individual merit, i.e., an objective assessment of the candidate's ability, performance, and potential—irrespective of sex, marital, or family status. With respect to recruitment and encouragement in the program, academic institutions and departments should be mindful of the need to redress past inequities and derelictions which have produced the present imbalance between the sexes in the economics profession. They should also be mindful of the positive contribution and enrichment of the educational process to be achieved through an increased participation of women. The profession can only stand to gain from the benefits of pluralist diversity and the recruitment of previously excluded, underutilized or unutilized talent.*

c) No employer, academic or otherwise, may discriminate against qualified women economists as candidates for positions, or for promotion or tenure, or for assignment of duties and responsibilities.

d) Salaries, fringe benefits, and facilities and resources for research for women shall be the same as those for men in the same position and rank.

e) There shall be no distinction between men and women as to what is considered to constitute full-time employment.

f) Employment may not be refused to any qualified economist on the grounds of family relationship to another employee.

ACTION RESPONSIBILITY: *In making appointment, retention, promotion, and tenure decisions, as well as in matters of compensation, fringe benefits, and prerequisites, women candidates shall be judged on their professional merit. The only relevant criteria shall be ability, potential, and performance. Care shall be exercised to avoid subtle as well as blatant, institutional as well as personal, forms of discrimination, especially where apparently innocuous regulations or policies impose differential burdens on women candidates. Institutional nepotism rules constitute the perhaps clearest departure from our central rule of judging candidates on their individual merit.*

II. RESOLVED that the American Economic Association establish a Committee on the Status of Women in the Economics Profession, numbering at least eight persons to be appointed by the President for a term of three years. The Association shall provide or seek funds for this Committee to investigate the extent of conformity to the principles stated above, and shall make recommendations for affirmative action. The Committee shall present an interim report at the next annual meeting. Its general findings, conclusions, and recommendations shall be published by the American Economic Association upon the Committee's request. When the Committee finds evidence of discrimination against women, it shall make this available for publication by the American Economic Association, and for the Association to present, as *amicus curiae*, in any complaint, remedial action, or suit.

ACTION RESPONSIBILITY: *Economists in educational institutions, government, and the private sector should help disseminate the findings of this Committee and its recommendations. The Committee wishes to assist in affirmative action programs and/or positive efforts to overcome role prejudice, but it also stands ready to receive complaints of discrimination. Rules to ensure proper procedures in handling such complaints are currently being formulated.*

III. RESOLVED that the American Economic Association shall encourage flexibility in providing opportunities for education and employment on a part-time basis for both men and women.

ACTION RESPONSIBILITY: *Making increasing provisions for education and employment on a part-time basis should be viewed as a technique for increasing individual opportunity and institutional flexibility. Such arrangements should not be allowed to serve as a cloak for evading compliance with affirmative action programs or for segregating women into underpaid and underprivileged professional categories.*

IV. RESOLVED that the American Economic Association appoint women economists as members of the editorial boards of its economic journals, that it actively encourage the appointment of women as program chairmen and participants at all future economic meetings, and that it urge companion associations to follow its example.

ACTION RESPONSIBILITY: *The Committee has written to the officers of regional and local economic associations as well as of allied social science associations, and to the editors of journals dealing with economics and economic affairs, urging that women economists be among those considered in selecting reviewers, referees, editorial board members, speakers, program arrangers, and so on. Those responsible for planning programs, conferences, commissioned articles or reports, and other opportunities for demonstrating professional achievement must remember that when women economists are less visible than men their competence or availability cannot be evaluated. The Committee has provided assistance in locating women economists and expects to continue to do so.*

IV. Conclusion

The recommended actions are only symbolic of needed changes in policy and attitude, if we are to correct the bizarre and irrational underrepresentation of women in the economics profession. The unemployment and underemployment of currently available talent, as well as the loss of potential talent, require positive, broad-gauged action to correct some structural imperfections in the professional labor market. To help eliminate these imperfections, one minimum step is continuing analytical study of the problem and regular, detailed reports to the Association and its members. In most instances—given the massive good will of our colleagues—the dissemination of information in itself will suffice to assure corrective action because, as Louis D. Brandeis once observed, "sunshine is the best disinfectant and electric light the most efficient policeman."[2] In other instances, to be sure, additional steps will be required to eliminate existing market imperfections and anomalies. This implies a continuing obligation of the Association and its members to consider alternative mechanisms, consistent with professional standards and ethics, to promote genuine equality of opportunity and meaningfully free access for all scholars and practitioners in the discipline of economics.

Committee Members

Carolyn Shaw Bell, Chair, Wellesley College
Martha O. Blaxall, Secretary, National Academy of Sciences
Walter Adams, Michigan State University
Francine Blau, Trinity College
Kenneth Boulding, University of Colorado
John Kenneth Galbraith, Harvard University
Collette Moser, Michigan State University
Barbara Reagan, Southern Methodist University
Myra Strober, Stanford University
Phyllis Wallace, Sloan School, Massachusetts Institute of Technology
Kenneth J. Arrow, Ex Officio, Harvard University

REFERENCES

M. B. Boyle, "Equal Opportunity for Women in Smart Business," *Harvard Bus. Rev.*, May-June 1973, *73*, 85–95.

L. W. Killough, letter to C. S. Bell, June 20, 1973.

A. Smith, *The Wealth of Nations,* New York 1937.

E. Tidball, "Perspective on Women and Affirmative Action," *Educational Rec.*, Spring 1973.

D. R. Witmer, letter to C. S. Bell, Apr. 15, 1973.

Amer. Econ. Rev. Proc., May 1972, *62*, 473–74.

Modern Language Association Newsletter, XXXIV, 9, 1972.

NOTES

Members with suggestions for the meetings in San Francisco, December 28–30, 1974, are invited to send them to Rendigs Fels, Secretary-Treasurer, American Economic Association, 1313 21st Avenue South, Nashville, Tennessee 37212. The Secretary-Treasurer will review the suggestions and forward them to the Program Chairman. Papers at sessions sponsored by the American Economic Association are ordinarily invited rather than contributed. To be considered, proposals for this Association must be received by the Secretary-Treasurer no later than February 1, 1974. Economists wishing to give papers on econometrics or economic theory may want to submit them to the Econometric Society, which meets with the Association and annually schedules a number of contributed papers.

The Executive Committee of the American Economic Association at its meeting on March 2, 1973 voted to increase dues 5 percent effective January 1, 1974, subject to reconsideration at its meeting December 27, 1973. This action was taken under the provision of the bylaws permitting the Executive Committee to increase dues proportionate to increases in relevant costs and prices.

American scholars scheduled to participate in international meetings abroad may, as in the past, receive a travel grant from the American Council of Learned Societies. Under the new procedures adopted by ACLS effective October 1973, the applicant must obtain a standard application form from the appropriate society and return the completed form to the society by the applicable deadline: February 15 (for meetings scheduled June to September), June 15 (for meetings scheduled October to January), or October 15 (for meetings scheduled February to May). Applicants will be carefully screened by the society. ACLS instructions state: "Awards are restricted to humanists and to those social scientists and legal scholars who are *strongly* oriented toward the humanities, who use humanistic methods in their research, and who will be participating in meetings concerned with the humanistic aspects of their disciplines." Inquiries should be addressed to Rendigs Fels, Secretary-Treasurer, American Economic Association, 1313 21st Avenue South, Nashville, Tennessee 37212.

Calls for Student Papers

The Conference on Research in Income and Wealth will include a student papers session in its meeting at the University of Michigan, May 15, 16, and 17, 1974. The focus of the meeting will be the distribution of income and wealth, broadly defined to include job perquisites and social benefits. A maximum of three papers by students will be selected for presentation and a discussant will be appointed for participation in the session. Any person currently working toward a degree at any level may offer a paper for consideration. The papers should be unpublished, no more than 30 pages, and not currently under consideration by a publisher. Interested students should send papers to the Conference Chairman, Dr. F. Thomas Juster, Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, Michigan 48106, by Feb 1, 1974.

Omicron Delta Epsilon is cosponsoring with Tax Analysts and Advocates a new annual essay contest on federal taxation named in honor of Paul H. Douglas, who has gained recognition both as a leading proponent of tax reform and as a distinguished public finance economist. The ODE-TAA annual contest is for the best undergraduate or graduate student essay of up to 5,000 words on any topic relevant to the U.S. federal tax system. The winning paper will be published in the Omicron Delta Epsilon journal, *The American Economist* and a cash award of $250 will be made. Two $100 awards will be made as second prize. Selection of winners will be made by Oct. 1, 1974. Essays should be sent to Tax Analysts and Advocates, 732 17th Street, Washington, D.C. 20006, by June 1, 1974.

The Southern Regional Science Association (*SRSA*) announces an award program for the best study submitted by undergraduate or graduate students, or those who have completed degree requirements in the academic year 1973–74. The author of the best study on a subject within the scope of regional science will receive an honorarium of $100 and be invited to present his or her paper at the annual *SRSA* meeting in Washington, D.C., April 4–5, 1974. All studies must be no longer than 15 typewritten pages, double spaced three copies submitted in edited form with footnotes at the end of the text. The name of the author his or her affiliation, and the title of the study should be on a separate cover sheet. The first text page should contain only the title and text. The cover page will be removed before the study is reviewed. All studies must be postmarked no later than Feb. 1, 1974 and sent to the Chairman of the *SRSA* Competition Committee: James M. Stepp, Department of Agricultural Economics, Clemson University, Clemson, South Carolina 29631.

The Journal of Law and Economics is in the process of collecting and cataloging the private papers of the late Henry C. Simons. They are seeking contributions to their collection in the form of manuscripts written by Simons and recollections or biographical materials by his associates and students as well as information concerning the location and contents of any other collections of Simons' papers. When completed, the collection will be available to scholars at the University of Chicago Law Library. Please send any contributions to R. H. Coase, *Journal of Law and Economics*, University of Chicago Law School, 1111 East 60th Street, Chicago, Illinois 60637.

The Office of Naval Research and the George Washington University, with the cooperation of the Air Force Office of Scientific Research and the Army Research Office, announce a Logistics Research Conference to be held at the George Washington University, Washington, D.C., May 8–10, 1974. Contributed papers are welcome. Abstracts and inquiries may be addressed to Ms. Henrietta Jones, Department of Operations Research, The George Washington University, Washington, D.C. 20006. Telephone 202+ 676-7504.

The Joint Committee on Eastern Europe of the American Council of Learned Societies and the Social Science Research Council wishes to draw attention to three of its grant programs: grants for postdoctoral research to a maximum of $3500; grants for study of East European languages to a maximum of $1000; grants ranging between $2000 and $5000 for the support of conferences. For details of eligibility and information which must be supplied in requesting application forms, write to Office of Fellowships and Grants, American Council of Learned Societies, 345 East 46 Street, New York, New York 10017.

The National Tax Association announces the 1973 award winners in the annual competition for outstanding doctoral dissertations in government finance and taxation. The $1,500 first prize award was won by John H. Bowman of Ohio State University (now at the Advisory Commission on Intergovernmental Relations), with his entry, "Cost and Benefit Spillouts as Factors Affecting Local Taxation for Public Schools." Honorable mention awards of $500 each were won by A. Mitchell Polinsky of Massachusetts Institute of Technology, (now at Harvard University), "Essays in Public Sector Economics: Central and Local" and Roger W. Schmenner of Yale University "City Taxes and Industry Location." The members of the 1973 Selection Committee were Professors Harvey E. Brazer, Arthur D. Lynn, Oliver Oldman, and James A. Papke. Information on the 1974 award competition may be obtained from Professor James A. Papke, Department of Economics, Krannert Graduate School of Industrial Administration, Purdue University, West Lafayette, Indiana 47906.

The Fourth World Congress of the International Economic Association will be held from August 19 to 24, 1974 in Budapest, Hungary, on the theme "Economic Integration: Worldwide, Regional, Sectoral." The program committee of 12 economists under the chairmanship of Fritz Machlup has invited 66 persons from 37 countries to fill places on the program. It provides for 12 main papers, 27 discussion papers, 10 opening statements for the discussions in separate working groups, 10 reports from chairmen of the working groups, and 7 chairmanships of plenary sessions. Many more, indeed several hundreds of those attending the Congress, will have an opportunity to participate in the discussions in the working groups, which will meet simultaneously in 5 half-day sessions.

The topics assigned to the working groups and their chairmen are: Measuring the Degree or Progress of Economic Integration (Eric Lundberg, Sweden); Sectoral Integration: Agriculture, Transport, Energy, Selected Industries (Joszef Pajestka, Poland); Industrial Policy: Location, Technology, Multinational Firms, Competition, and Integration of Product Markets (Harry G. Johnson, U.K.); Migration and Integration of Labor Markets (Orlando d'Alauro, Italy); International Capital Movements and Integration of Capital Markets (Herbert Giersch, Western Germany); Monetary and Fiscal Integration (Raymond Barre, France); Socio-Political and Institutional Aspects of Integration (Shigeto Tsuru, Japan); Integration of Less Developed Areas and of Areas on Different Levels of Development (I. G. Patel, India); Integration by Market Forces and through Planning (Bela Czikos-Nagy, Hungary); World Trade and Intraregional Trade: Trends and Structural Changes (P. J. Verdoorn, Netherlands). The following economists from the United States will present main papers, discussion papers, or opening statements: Bela Balassa, Richard N. Cooper, Gottfried Haberler, Peter Kenen, and Marina v. N. Whitman. The complete program will be published in the March 1974 issue of this *Review*.

Attendance is open to all members of the American Economic Association and of other national associations belonging to the I.E.A. The Hungarian arrangements committee has reserved over 2000 beds in Budapest. It is hoped that several hundred economists from the United States will attend. Registration fee at the Congress will be $50—reduced to $30 for early reservation up to March 31, and $40 from April 1 to July 31. Fees for students are one-half, and fees for members of the family accompanying a participant are $5 each. Forms for registration will be distributed through universities, research organizations, the secretariat of the AEA, and by direct mail from the Hungarian Economic Association, 1370 Budapest, Pf. 544, Hungary.

The International Economic Association will hold a workshop on "Population and Development," in cooperation with Bosphorus University, Istanbul, from September 2 to 14, 1974. The workshop will consider interrelations between economic and demographic factors in development. Attendance will be limited to 25 economists with experience in development programs. They need not have prior specialized experience in the subject of population. Expert discussion leaders will provide guidance. Preference will be given to senior and other responsible officials of government but academic economists will also be considered. All proceedings will be in English.

Persons wishing to be considered for invitation may apply by letter to the organizing committee. Please describe your relationship to development administration and policy, and include a curriculum vitae. Fluency in English is required. Participants attending the entire workshop will receive transportation, maintenance, and a modest per diem. Write to Dr. Paul Demeny, Director, The Population Council, 245 Park Avenue, New York, New York 10017, U.S.A. Other committee members are Professor Emre Gonensay and Dr. Frederic C. Shorter. Inquiries will be accepted up to March 15, and invitations will be extended by April 15, 1974.

The number of confirmed appointments through the National Registry for Economics, which was established in 1966, exceeds 1,200. Mrs. Theresa R. Scholl has been its manager since its inception. The National Registry is part of the nationwide professional placement network of the U.S. Employment Service, which is now celebrating its fortieth anniversary. For further details, see the section below on Employment Services.

Papers are solicited for the sixth Amos Tuck School Seminar on Problems of Regulation and Public Utilities, sponsored by the American Telephone and Telegraph Company, to be held in late August 1974 at Dartmouth College. The seminar will be attended by 30 conferees chosen from academic institutions throughout the country. It consists of 7 or 8 sessions over a 4-day period. The sessions provide a forum for the presentation and discussion of new ideas and innovative work in analyzing regulated firms, regulatory processes, and problems of public welfare.

We invite those with an interest in this area to suggest names of possible speakers or to apply themselves. The aim of the seminar is to introduce the research both of young scholars beginning their careers and of scholars who have only recently turned their attention to public utility problems. Therefore, suggestions of speakers must be limited to scholars in these categories. Those chosen to present papers will receive support for research and writing, in addition to travel and living expenses at the seminar. Please send suggestions to the coordinator of the seminar before February 23, 1974. Include two copies of each of the following: biographical information; samples of written work; and a proposal to be developed for presentation.

This is not a request for conferee applications. Application forms and further information will be mailed to institutions in March 1974. The director of the seminar is Professor Duncan M. Holthausen, Jr. Please direct all suggestions and correspondence to: Coordinator, Seminar on Problems of Regulation and Public Utilities, The Amos Tuck School of Business Administration, Dartmouth College, Hanover, New Hampshire 03755.

Deaths

Richard C. Bernhard, professor of economics, University of Utah, June 19, 1973.

David M. Harrison, professor emeritus, Ohio State University, June 17, 1973.

John W. Leonard, professor of economics and dean of School of Business, Southern Illinois University, Mar. 18, 1973.

Karl Pribram, professor emeritus, Frankfurt University; economist, retired, U.S. government; Washington, D.C., July 14, 1973.

Retirements

Claude A. Campbell, professor emeritus of finance, Florida State University, June 1973.

Troy J. Cauley, professor of economics, Indiana University, June 15, 1973.

J. Herschel Coffee, department of economics, West Texas State University, Aug. 31, 1973.

George E. Hunsberger, department of economics, College of Business Administration, University of Arkansas, June 30, 1973.

Janis Labsvirs, Indiana State University, professor of economics, June 30, 1973.

Charles N. Lanier, professor of economics and business administration, University of Delaware, Sept. 1973.

Herman E. Michl, professor of economics and business administration, University of Delaware, Sept. 1973.

Frances Simmons, Western Kentucky University, June 1973.

Philip E. Taylor, professor and former head, department of economics, University of Connecticut, Sept. 1973.

Visiting Foreign Scholars

Geoffrey P. Braae, University of Auckland: visiting professor, department of economics and commerce, Simon Fraser University, Sept. 1, 1973.

Yakar Kannai, Hebrew University, Jerusalem: visiting professor, department of mathematics, Center for Mathematical Studies in Economics and Management Science, Northwestern University, Sept. 1973-June 1974.

Edi Karni, Tel Aviv University: visiting assistant professor, Ohio State University, summer quarter 1973.

John R. Moroney, Tulane University: visiting professor, department of economics and commerce, Simon Fraser University, May 1, 1974.

Aba Schwartz, Tel Aviv University: visiting associate professor, Ohio State University, autumn quarter 1973.

Turgut Var, University of Kansas: visiting associate professor, department of economics and commerce, Simon Fraser University, Sept. 1, 1973.

Promotions

Robert F. Adams: professor of economics, University of California, Santa Cruz, July 1, 1973.

Theodore E. Allison: assistant secretary, Office of the Secretary, Board of Governors of the Federal Reserve System, June 10, 1973.

B. Wylie Anderson: associate professor of economics, University of Northern Iowa, Sept. 1973.

Ahmed N. A. Azim: associate professor, faculty of business, University of Calgary, July 1, 1973.

Forrest C. Blodgett: associate professor of economics, Linfield College.

Edward G. Boehne: senior vice president, Federal Reserve Bank of Philadelphia, Mar. 1, 1973.

James M. Boughton: associate professor of economics, Indiana University, July 1, 1973.

Paul L. Burgess: associate professor, Bureau of Business and Economic Research, Arizona State University, Sept. 1973.

Kenneth T. Cann: professor, department of economics, Western Kentucky University, Aug. 1973.

Charles Cathcart: assistant professor of economics, Pennsylvania State University, July 1, 1973.

Venkareddy Chennareddy: senior economist, Economic Research and Planning, Southern Research Institute.

Whewon Cho: professor of economics, Tennessee Technological University.

Ronald W. Chorba: associate professor, faculty of business, University of Calgary, July 1, 1973.

Caroline Clotfelter: assistant professor of economics, Mercer University, Sept. 1973.

Richard G. Davis: vice president, Research and Statistics Function, Federal Reserve Bank of New York.

Donn A. Derr: associate professor, Rutgers University, July 1, 1973.

Barry L. Duman: associate professor of economics, West Texas State University, Sept. 1, 1973.

Irwin Feller: professor of economics, Pennsylvania State University, July 1, 1973.

Ismail A. Ghazalah: associate professor, department of economics, Ohio University, Sept. 1, 1973.

W. Lee Hoskins: vice president-research, Federal Reserve Bank of Philadelphia, Mar. 1, 1973.

Katherine Hsiao: professor of economics, Indiana State University, Aug. 23, 1973.

David N. Hyman: associate professor of economics, North Carolina State University, July 1, 1973.

John E. Ikerd: extension associate professor of economics, North Carolina State University, July 1, 1973.

Ira P. Kaminow: economic adviser, Federal Reserve Bank of Philadelphia, Mar. 1, 1973.

Jerry L. Kingston: associate professor, Bureau of Business and Economic Research, Arizona State University, Sept. 1973.

Michael A. Klein: associate professor of economics, Indiana University, July 1, 1973.

V. N. Krishnan: associate professor, department of economics, Bowling Green State University, Sept. 1, 1973.

Robert J. Lawrence: associate adviser, Division of Research and Statistics, Board of Governors of the Federal Reserve System, June 10, 1973.

Robert L. Lawson: professor of economics, Ball State University, Sept. 1973.

Wayne A. Long: associate professor, faculty of business, University of Calgary, July 1, 1973.

James B. McDonald: associate professor of economics, Brigham Young University, Sept. 1973.

Charles E. McLure, Jr.: professor in economics and finance, Rice University, July 1973.

Paul Meek: monetary adviser, Open Market Operations and Treasury Issues Function, Federal Reserve Bank of New York.

Jacob B. Michaelsen: professor of economics, University of California, Santa Cruz, July 1, 1973.

Donald J. Mullineaux: senior economist, Federal Reserve Bank of Philadelphia, Mar. 1, 1973.

John A. Naylor: associate professor, department of economics, College of Wooster, Sept. 1, 1973.

Jon P. Nelson: associate professor of economics, Pennsylvania State University, July 1, 1973.

E. C. Pasour, Jr.: professor of economics, North Carolina State University, July 1, 1973.

Nicholas S. Perna: chief, Business Conditions Division, Federal Reserve Bank of New York.

A. Marshall Puckett: vice president, Accounting Control Function, Federal Reserve Bank of New York.

John H. Riew: professor of economics, Pennsylvania State University, July 1, 1973.

Richard L. Ruth: professor of economics, Northeastern Illinois University, July 1, 1973.

Gerald W. Sazama: associate professor of economics, University of Connecticut, Oct. 1973.

Vishwa S. Shukla: associate professor, department of economics, Ohio University, Sept. 1, 1973.

Nancy D. Sidhu: associate professor of economics, Northeastern Illinois University, Sept. 1973.

James D. Smith: professor of economics, Pennsylvania State University, July 1, 1973.

Ronald Soligo: professor of economics, Rice University, July 1973.

W. E. Spellman: associate professor of economics, Coe College, Sept. 1, 1973.

Gary H. Stern: manager, domestic research department, Federal Reserve Bank of New York.

Edmund G. Sugars: professor, faculty of business, University of Calgary, July 1, 1973.

Rudolf Thunberg: assistant vice president, Research and Statistics Function, Federal Reserve Bank of New York.

George M. von Furstenberg: professor of economics, Indiana University, July 1, 1973.

Howard M. Wachtel: associate professor, department of economics, American University, Sept. 1, 1973.

Ronald D. Watson: senior economist, Federal Reserve Bank of Philadelphia, Mar. 1, 1973.

William W. Wiles: assistant director, Division of Supervision and Regulation, Board of Governors of the Federal Reserve System, June 10, 1973.

James E. Zinser: associate professor of economics, Oberlin College.

Administrative Appointments

B. Wylie Anderson: head, department of economics, University of Northern Iowa, Sept. 1973.

Earl Baggerly: dean, Night School, Mercer University, July 1973.

Peter S. Barth, U.S. Department of Labor: head, department of economics, University of Connecticut, Sept. 1, 1973.

Raymond R. Beneke: chairman, department of economics, Iowa State University, Apr. 16, 1973.

Eric Brucker: chairman, department of economics, University of Delaware, Sept. 1973.

Sherrill Cleland, Kalamazoo College: president, Marietta College and professor of economics, July 1, 1973.

Robert F. Cook: chairman, department of economics, Indiana State University, Aug. 23, 1973.

Ervin J. Doak: chairman, department of economics, Saint Mary's University, Halifax, July 1, 1973.

Barry L. Duman: head, department of economics, West Texas State University, Sept. 1, 1973.

John W. Hannaford: associate dean, College of Business, Ball State University, Sept. 1972.

Robert T. Holland: co-chairman, Graduate Division, and co-chairman, Division of Business Administration, Woodbury College, May 1972.

Robert R. Jost: chairman, department of economics, Ball State University, Sept. 1972.

James A. Kane, University of Massachusetts: assistant secretary of Manpower Affairs, Commonwealth of Massachusetts, May 1973.

Richard D. Karfunkle, Laird, Bissell & Meeds, Inc.: vice president and chief economist, Abraham & Co., Inc., Delaware.

Hirschel Kasper: chairman, department of economics, Oberlin College.

Pauline W. Kopecky: director, University Affirmative Action Program, Oklahoma State University, Aug. 15, 1973.

Jacob B. Michaelsen: chairman, Board of Studies in Economics, University of California, Santa Cruz, July 1973.

Edward H. Murphy: assistant to the dean, faculty of Business, University of Calgary, 1973–74.

John C. Narver: chairman, department of marketing, transportation, and international business, University of Washington, Sept. 1974.

Leo J. Navin: chairman, department of economics, Bowling Green State University, Sept. 1973.

Stephen G. Peitchinis: dean, faculty of business, University of Calgary, Apr. 1, 1973.

Richard L. Ruth, Saint Mary's University, Halifax: chairman department of economics, Northeastern Illinois University, July 1, 1973.

James Satterfield: director of Admissions, Mercer University, spring 1973.

A. Ross Shepherd, assistant provost, University of Missouri-Kansas City, June 1, 1973.

David J. Smyth: chairman, department of economics, Claremont Graduate School, July 1, 1973.

Gary C. Taylor, National Water Commission: assistant deputy administrator, Economic Research Service, USDA, July 1973.

Walter J. Wadycki: assistant vice chancellor for academic affairs, University of Illinois at Chicago Circle, Aug. 16, 1973.

Frank E. Wagner: chairman, department of economics, University of Missouri-Kansas City, June 1, 1973.

James H. Weaver: chairman, department of economics, American University, Sept. 1, 1973.

Elmus R. Wicker: chairman, department of economics, Indiana University, July 1, 1973.

Robert Willard: chairman, Division of Business and Economics, Mercer University, July 1973.

New Appointments

Fred J. Abraham: assistant professor of economics, University of Northern Iowa, Sept. 1973.

Adebayo Akerele: lecturer, department of economics, Rutgers-The State University, July 1973.

Robert J. Anderson: associate professor of economics, director, Center for Study of Environmental Policy, Pennsylvania State University, July 1, 1973.

V. Balachandran, Carnegie-Mellon University: associate professor of decision sciences, department of managerial economics and decision sciences, Northwestern University, Sept. 1973.

M. Alan Baughcum: instructor of economics, North Carolina State University, Aug. 1973.

David C. Beek: economist, Balance of Payments Division, Federal Reserve Bank of New York.

Thomas E. Borcherding, Virginia Polytechnic Institute and State University: associate professor, department of economics and commerce, Simon Fraser University, Sept. 1, 1973.

Leigh B. Boske, Stockton State College: senior economist, Wisconsin Department of Transportation, May 1973.

William W. Brooks: instructor of economics, North Carolina State University, Aug. 1973.

Leonard R. Burgess: visiting professor of management, Temple University, Sept. 1973.

Richard Butler, Massachusetts Institute of Technol-

ogy: assistant professor of economics, Rice University, July 1, 1973.

C. Stuart Callison, Cornell University: assistant professor of economics, Ohio University, Sept. 1, 1973.

Geoffrey Carliner, University of Wisconsin, Madison: assistant professor, department of economics, Oberlin College, Sept. 1973.

Pao Lun Cheng, University of Massachusetts: professor, department of economics and commerce, Simon Fraser University, Jan. 1, 1974.

Peter K. Clark: visiting research fellow, National Bureau of Economic Research, 1973–74.

Benjamin I. Cohen: research associate, National Bureau of Economic Research, July 1, 1973.

A. Edward Day: assistant professor of economics, University of Nebraska-Lincoln, Sept. 1973.

Paul DeRosa: economist, Money and Finance Division, Federal Reserve Bank of New York.

James E. Easley, Jr.: instructor of economics, North Carolina State University, Sept. 1973.

Michael G. Ellis: assistant professor, department of economics, New Mexico State University, Aug. 1973.

Thomas J. Espenshade, Bowdoin College: assistant professor of economics and research associate of Population and Manpower Research Center, Florida State University, Sept. 1973.

Mary L. Eysenbach, University of Washington: assistant professor of economics, University of Utah.

Douglas Fisher: visiting professor of economics, Claremont Graduate School, Sept. 1973.

David T. Geithman: associate professor of economics, Russell Sage College, Sept. 1, 1973.

Philip M. Ginsberg: assistant professor, department of economics, Rutgers-The State University, July 1973.

Gerald Glyde: assistant professor of economics, Pennsylvania State University, Sept. 1, 1973.

Clive W. J. Granger, University of Nottingham, England: professor of economics, University of California, San Diego, July 1, 1973.

Michael J. Greenwood: associate professor, Bureau of Business and Economic Research, Arizona State University, June 1973.

Theodore Groves, University of Wisconsin: associate professor of managerial economics, department of managerial economics and decision sciences, Northwestern University, Sept. 1973.

Daniel S. Hamermesh, Princeton University: associate professor, department of economics, Michigan Stata University, Sept. 1973.

Stephen K. Happel: instructor of economics, North Carolina State University, Aug. 1973.

Kurt Hausafus, University of Illinois, Chicago Circle: assistant professor, department of economics, Oberlin College, Sept. 1973.

Dennis R. Heffley: assistant professor of economics, University of Connecticut, Sept. 1, 1973.

Anthony F. Herbst: associate professor, faculty of business, University of Calgary, July 1, 1973.

Dana C. Hewins, University of Illinois: assistant professor of economics, Ohio University, Sept. 1, 1973.

Mary C. Hilliard: instructor of economics, North Carolina State University, Aug. 1973.

Mary T. Holcomb: instructor of economics, North Carolina State University, Aug. 1973.

L. Dwight Israelsen, Massachusetts Institute of Technology: assistant professor of economics, Brigham Young University, Sept. 1973.

Suresh Jain, Yale University: assistant professor of decision sciences, department of managerial economics and decision sciences, Northwestern University, Sept. 1973.

William G. Johnson, Rutgers University: assistant professor of economics, Syracuse University, Sept. 1, 1973.

Sergei Kasakow: visiting assistant professor, Bureau of Business and Economic Research, Arizona State University, Sept. 1973.

Donald W. Katzner, University of Waterloo: lecturer in economics, University of California, San Diego, July 1, 1973.

Herbert M. Kaufman: assistant professor, Bureau of Business and Economic Research, Arizona State University, Sept. 1973.

Christopher King: instructor of economics, University of Utah.

Jack L. Knetsch: professor, department of economics and commerce, Simon Fraser University, May 1, 1974.

Conway L. Lackman: assistant professor, department of economics, Rutgers-The State University, July 1973.

John Landon, Case-Western University: associate professor of economics, University of Delaware, Sept. 1973.

Stephen B. Levinson: instructor, department of economics, Rutgers-The State University, July 1973.

Kenneth Lewis, Boston College: associate professor of economics, University of Delaware, Sept. 1973.

Stephen E. Lile: associate professor of economics, Western Kentucky University, Aug. 1973.

Cliff Lloyd, University of Iowa: professor, department of economics and commerce, Simon Fraser University, Sept. 1, 1973.

Donald C. MacCharles: department of economics, McMaster University, July 1973.

William A. McEachern: assistant professor of economics, University of Connecticut, Sept. 1, 1973.

James A. McMahon, I.C.A.O., Montreal: industry economist, Board of International Affairs, Civil Aeronautics Board, Sept. 1973.

J. C. McManus, Carleton University: associate professor, department of economics, McMaster University, 1973–74.

W. Edward McMullan: assistant professor, faculty of business, University of Calgary, July 1, 1973.

Sherman J. Maisel: senior research staff, National Bureau of Economic Research, June 1, 1973.

Vincent J. Malanga, Fordham University: economist, Federal Reserve Bank of New York, Sept. 1, 1973.

John P. Matilla, Ohio State University: assistant

professor of economics, Iowa State University, Sept. 1, 1973.

Eileen Mauskopf: lecturer, department of economics, Rutgers-The State University, July 1973.

Robert H. Miles: instructor of economics, North Carolina State University, Aug. 1973.

Paul A. Montavon: economic advisor, U.S. Permanent Mission to the Organization of American States, July 23, 1973.

Francis Mulvey: assistant professor, department of economics, Bowling Green State University, Sept. 1973.

Dick K. Nanto, Harvard University: assistant professor of economics, Brigham Young University, Sept. 1973.

Philip Nelson: research associate, National Bureau of Economic Research, July 1, 1973.

Norman P. Obst, University of Washington: assistant professor, department of economics, Michigan State University, Sept. 1973.

Robert T. Owen: assistant professor, faculty of business, University of Calgary, July 1, 1973.

Avinoam Perry, Indiana University: assistant professor of decision sciences, department of managerial economics and decision sciences, Northwestern University, June 1973.

James R. Prescott, Temple University: professor of economics, Iowa State University, Sept. 1, 1973.

David Puryear, Princeton University: assistant professor of economics, Syracuse University, Sept. 1, 1973.

Donald Ratajczak, University of California, Los Angeles: associate professor of economics and director, Georgia State University Economic Forecasting Project, Georgia State University, July 1, 1973.

Roger Riefler: associate professor of economics, University of Nebraska-Lincoln, Sept. 1973.

R. Lynn Rittenoure, University of Texas-Austin: assistant professor of economics and operations research, Navy Management Systems Center, Naval Postgraduate School, Sept. 1, 1973.

Roy Rotheim: assistant professor, department of economics, Bowling Green State University, Sept. 1973.

R. Robert Russell, University of California, Santa Barbara: associate professor of economics, University of California, San Diego, 1972–73.

Daniel H. Saks: visiting research fellow, National Bureau of Economic Research, 1973–74.

Bertram Schoner, University of Iowa: professor, department of economics and commerce, Simon Fraser University, May 1, 1973.

Robert A. Schultz: associate professor, faculty of business, University of Calgary, July 1, 1973.

Suresh P. Sethi, Rice University: visiting assistant professor, faculty of management studies, University of Toronto, July 1973.

Berndt A. Sigloch: instructor, department of economics and commerce, Simon Graser University, Sept. 1, 1973.

Irwin H. Silberman, New York University: associate professor, Queens College, City University of New York, 1973–74.

Ron Stanfield: assistant research environmental economist, Center for Marine Affairs, Scripps Institution of Oceanography, Aug. 1, 1973.

Wayne Stevenson, Middlebury College: assistant professor of economics, University of Utah.

Edward Stohr, University of California, Berkeley: assistant professor of decision sciences, department of managerial economics and decision sciences, Northwestern University, Sept. 1973.

Arie Tamir, Case-Western Reserve University: visiting assistant professor of decision sciences, department of managerial and decision sciences, Northwestern University, Sept. 1973.

Marie C. Thursby: visiting assistant professor of economics, North Carolina State University, Aug. 1973.

William P. Travis: professor of economics, Indiana University, Jan. 1974.

William H. Wallace: professor of economics and assistant department head for management, North Carolina State University, July 1973.

Yoram Weiss: research associate, National Bureau of Economic Research, Sept. 1, 1973.

J. D. Welland: department of economics, McMaster University, Jan. 1, 1974.

Roy E. Welsch: senior research associate, National Bureau of Economic Research, Computer Research Center, June 1, 1973.

John Wenninger: economist, Market Statistics Division, Federal Reserve Bank of New York.

Michael R. Wickens: adjunct associate professor, department of economics, Rutgers-The State University, July 1973.

W. Philip Windham: special lecturer of economics, North Carolina State University, Aug. 1973.

Christine Wollan: economist, European department, International Monetary Fund, May 1973.

Jong Keun You: lecturer, department of economics, Rutgers-The State University, July 1973.

Leaves for Special Appointments

Ghassan M. Arnaoot, West Texas State University: International Monetary Fund, Aug. 1, 1973.

Robert E. Berney, Washington State University: senior academic resident in public finance, Advisory Commission on Intergovernmental Relations, Aug. 1973.

Bernard H. Booms, Pennsylvania State University: economist in residence, City Manager's Office, Tacoma, Washington, July 1973–Aug. 1974.

James M. Boughton, Indiana University: administrator, Monetary Division, economics and statistics department, Organization for Economic Cooperation and Development, Sept. 1973.

Daniel W. Bromley, University of Wisconsin, Madison: economist, Office of the Secretary, U.S. Department of the Interior, Washington, D.C., Sept. 1973–Aug. 1974.

Leland S. Case, Brigham Young University: senior research economist, Association of American Railroads, Sept. 1973.

Edwin G. Charle, Ohio University: Fulbright Fellow, Cuttington College, Monrovia, Liberia, July 1, 1973.

Charles R. Chittle, Bowling Green State University: Institute for World Economics, Kiel University, Germany, June 1973–Aug. 1974.

Steven R. Cox, Arizona State University: Federal Trade Commission, July 1973.

Thomas F. Dernburg, Oberlin College: fiscal economist, International Monetary Fund, Feb. 1, 1973.

William G. Dewald, Ohio State University: special assistant for economic analysis and research, International Labor Bureau, U.S. Department of Labor, Washington, D.C., winter and spring quarters 1974.

Marvin R. Jackson, Arizona State University: Senior Fulbright-Hayes Program, Romania.

Roland H. Koller II, Brigham Young University: economist, Federal Trade Commission, July 1973.

Arlyn J. Larson, Arizona State University: State of Arizona, 1973–74.

Ben N. Matta, New Mexico State University: U.S. Department of Housing and Urban Development, Fort Worth, Texas, summer 1973.

Charles W. Meyer, Iowa State University: Long Range Research Branch, Social Security Administration, Washington, D.C., Sept. 1, 1973–Aug. 31, 1974.

John C. Narver, University of Washington: visiting professor, Norwegian College of Economics and Business Administration, Sept.–Dec. 1973.

James R. Ostas, Bowling Green State University: Federal Reserve Bank of Boston, June 1973–Aug. 1975.

Refugio I. Rochin, University of California, Davis: program assistant for rural development, The Ford Foundation, Bogota, Colombia, July 1, 1973–Sept. 1975.

Richard L. Schmalensee, University of California, San Diego: visiting assistant professor, Institute of Economics, Université Catholique de Louvain, Belgium, 1973–74.

David Segal, Oberlin College: visiting scholar, Resources for the Future, 1973–74.

S. W. Son, Slippery Rock State College: senior staff economist, Council of Economic Advisers, Washington, D.C. July 1973.

Dennis R. Starleaf, Iowa State University: visiting professor, economic section, Board of Governors of the Federal Reserve System, Washington, D.C., Sept. 1, 1973–Aug. 31, 1974.

Daniel B. Suits, University of California, Santa Cruz: visiting professor of economics, University of Hawaii, 1973–74.

P.A.V.B. Swamy, Ohio State University: consultant, Board of Governors of the Federal Reserve System, Washington, D.C., autumn quarter 1973.

George M. von Furstenberg, Indiana University: senior economist, Council of Economic Advisers, Washington, D.C., Sept. 1973.

Percy D. Warner III, University of Connecticut: Fulbright Fellow, Brazil, Sept. 1, 1973–July 31, 1974.

John C. Weicher, Ohio State University: U.S. Department of Housing and Urban Development, Washington, D.C.

Resignations

Terrell D. Blodgett, Southwest Texas State University, June 1973.

Eliyahu Borukhov, Ohio State University: Tel Aviv University, Sept. 1973.

J. Hayden Boyd, Ohio State University: Economics Division of Motor Vehicle Manufacturers Association, Sept. 1973.

Thomas F. Cargill, Purdue University: University of Nevada-Reno, Aug. 1973.

Richard J. Cebula, Ohio University: Emory University, June 30, 1973.

Ibrahim Eris, Rice University: University of São Paulo, July 1973.

Dieter Nowak, University of Calgary, June 30, 1973.

Philip D. Olson, University of Calgary, June 30, 1973.

David L. Paden, IU International Corporation: K. W. Tunnell Company, Inc., Aug. 1, 1973.

Charles E. Seagrave, Rice University: H.E.W., Washington, D.C., Dec. 1972.

Theodore L. Vaugh, Marietta College: Murray Fabrics, Inc, Kentucky, May 1, 1973.

NOTE TO DEPARTMENTAL SECRETARIES AND EXECUTIVE OFFICERS

When sending information to the *Review* for inclusion in the Notes Section, please use the following style:

A. Please use the following categories:

1—Deaths
2—Retirements
3—Foreign Scholars (visiting the USA or Canada)
4—Promotions
5—Administrative Appointments
6—New Appointments
7—Leaves for Special Appointments (NOT Sabbaticals)
8—Resignations
9—Miscellaneous

B. Please give the name of the individual (SMITH, John W.), his present place of employment or enrollment:

his new title (if any), his next place of employment (if known or if changed), and the date at which the change will occur.

C. Type each item on a separate 3x5 card, and please do not send public relations releases.

D. The closing dates for each issue are as follows: *March,* November 1; *June,* February 1; *September,* May 1; *December,* August 1.

This announcement supersedes and replaces a letter which was sent annually from the managing editor's office. All items and information should be sent to the Assistant Editor, *American Economic Review,* Box Q, Brown University, Providence, Rhode Island 02912.

SEVENTIETH LIST OF DOCTORAL DISSERTATIONS IN POLITICAL ECONOMY IN AMERICAN UNIVERSITIES AND COLLEGES

The present list specifies doctoral degrees conferred during the academic year terminating June 1973. Abstracts of many of the dissertations are supplied.

General Economics; including Economic Theory, History of Thought, Methodology, Economic History, and Economic Systems

EILEEN APPLEBAUM, Ph.D. Pennsylvania 1972. Toward a nonmarginalist approach to price formation and income distribution.

ELIZABETH E. BAILEY, Ph.D. Princeton 1972. Economic theory of regulatory constraint.

Part I presents a general framework that can accommodate many interesting models of regulatory constraint. Part II treats the most important interpretation of the general structure, namely, the Averch-Johnson model of rate-of-return regulation. Some ways of altering the model to include issues such as regulatory lag, peak-load pricing, and a rising cost of capital are also discussed. Wherever possible, the properties of the regulatory models are described both graphically and mathematically.

ROGER L. BECK, Ph.D. Chicago 1972. Patents, property rights, and social welfare: Search for a restricted optimum.

Several models are used to analyze the effect of structural features of the patent system on the system's objective: maximization of social profit on innovations. One implication is selected for testing—that social profit will be the most severely reduced by patenting of substitutes where profits on innovation are greatest. Data covering a period of several years in three industries are consistent with the hypothesis.

ALAN R. BECKENSTEIN, Ph.D. Michigan 1972. An optimization approach for evaluating multiplant scale economies.

ROBIN W. BOADWAY, Ph.D. Queen's 1973. Some aspects of the theory of externalities.

JOHN P. BONIN, Ph.D. Rochester 1973. Target specification, managerial rewards, and the socialist firm.

Optimal policies for a socialist firm are investigated in both static and dynamic models by considering the effects of various target and managerial reward specifications on quantity decisions. When targets are misspecified, the information generated from observing the firm's behavior is discussed. Misspecification involves targets which do not satisfy the firm's production function or input targets which are not cost minimizing for given factor prices. Finally, a preliminary investigation into intertemporal target variation is undertaken.

MARCEL BOYER, Ph.D. Carnegie-Mellon 1973. Essays in optimal growth theory: A dynamic programming approach.

JAMES B. BRADFIELD, Ph.D. Rochester 1973. A dynamic programming model of the stock market specialist.

Our specialist makes a market by announcing prices periodically throughout the trading day. He guarantees these prices by trading for his own account. His overnight inventory positions incur holding costs. The specialist maximizes the discounted expected value of his own net trading profits over a finite horizon of trading days. The optimal announced price depends on the time of day, the level of his inventory, and his exclusive knowledge of the "book" of limited-price orders.

DENNIS NAI-WEI CHAO, Ph.D. California (Santa Barbara) 1972. Consumer behavior, activity chains and socioeconomic model of family formation.

GEORGE SAW-HO CHUNG, Ph.D. Southern California 1973. Differential game theory approach to modeling dynamic imperfect market processes.

NINA W. CORNELL, Ph.D. Illinois (Urbana) 1972. The role of the nobility in agricultural change in Russia during the reign of Catherine II.

JAMES F. CROOK, Ph.D. Virginia Polytechnic Institute 1973. Separate vs. concurrent decision procedures for public goods provision under majority rule.

This study attempts to develop a theoretical rationale for the choice between separate and concurrent decision procedures by voting groups. Concurrent procedures involve simultaneous decisions on the outputs of many public goods while separate procedures determine outputs of public goods singly or in large subsets. General purpose legislatures are examples of the former and referenda or special purpose legislative boards are examples of the latter.

DAVID DAYAN, Ph.D. Princeton 1973. Vertical intergration and monopoly regulation.

This thesis deals with the motivation of the regulated firm to exercise vertical control, the effect of

vertical control on the firm's behavior, and, conversely, the impact of various regulatory policies on the behavior of the vertically integrated firm. Several models of the single stage and multistage firm under constraints are developed and analyzed using non-linear optimization techniques. The models are compared and assessed in terms of their implications for economic theory, practice, and public policy.

FLOYD R. DILL, Ph.D. Cornell 1973. The institutional possibilities and limitations of economic progress in antebellum Georgia: A social systems approach.

RICHARD C. EDWARDS, Ph.D. Harvard 1972. Alienation and inequality: Capitalist relations of production in bureaucratic enterprises.

This thesis investigates the internal structure of modern enterprises. Contradictions growing out of earlier systems of power have led capitalists to attempt to maintain control by embedding power relations in the organizational or technical structure of work. An empirical analysis based on data for 455 workers shows that the firm's incentives are significantly geared to control, and that firms highly reward control-related "noncognitive" worker characteristics. A final analysis (N=217) shows that similar noncognitive traits are importantly related to success in high school.

JOHN P. FARRELL, Ph.D. Wisconsin (Madison) 1973. Aspects of central banking in Poland, 1950–1970.

ANN N. P. FISHER, Ph.D. Connecticut 1973. Durable goods and consumer choice.

Demand decisions for ordinary continuously divisible commodities and durables are studied, where discrete durables may be purchased on installment plans. Via a two-stage process, the consumer is hypothesized to rank all affordable bundles of (dated) durables (with their payment plans), each matched with its optimal bundle of divisible goods. The effect on demand for ordinary goods resulting from price, income, or payment plan changes can be predicted only if the chosen durable bundle remains the same.

ARTHUR D. FRAAS, Ph.D. California (Berkeley) 1972. The impact of the policies of the second bank of the United Slates on the economy of the Ohio Valley: 1814–1829.

LEWIS FREIBERG, JR., Ph.D. Kentucky 1972. Political coalitions and the allocation of resources.

This dissertation builds an axiomatic model of direct democracy to examine the efficiency of the political allocation of resources. Pareto optimality, operationalized through Scitovsky's double compensation criteria, is used to examine the welfare effects of appending a bribery process to the model. Several theorems are then derived concerning the type, direction, size, and welfare effects of resulting bribes. With these theorems, one can predict the outcome and Pareto optimality of the resulting political allocation of resources.

GERARD O. GAUDET, Ph.D. Pennsylvania 1972. Imperfections in the market for capital goods: Their effects on optimal firm size and on the relative efficiency of monetary and fiscal policy.

DAVID E. R. GAY, Ph.D. Texas A&M 1973. Capital and the production process: A critical evaluation of the Böhm-Bawerk—Clark debate and its relation to current capital theory.

DONALD W. GREEN, Ph.D. California (Berkeley) 1972. Industrialization and the engineering ascendancy: A comparative study of American and Russian elites, 1870–1920.

SURINDER S. GUJRAL, Ph.D. Georgetown 1973. Population growth and issues in Indian economic history, 1851–1901.

The study provides a complete series on India's population, by provinces and states, and labor force estimates for the decades between 1851 and 1901. The analysis of the data questions the widely held contention that India experienced a rapid increase in population prior to 1872. The treatment of the period 1872 to 1921 as homogeneous with respect to population growth is questioned. The labor force estimates represent a marked departure from previous analysis of the census data on occupational distribution.

DOUGLAS R. HALE, Ph.D. Kansas 1972. Imperfect competition and general equilibrium.

This dissertation is concerned with establishing the existence of equilibrium for monopolistically competitive economies. Following Negishi, each monopolistic competitor is viewed as maximizing his profit relative to a perceived demand. The assumptions applying to individuals, competitive firms, and the resource endowment are explained. Sufficient conditions for the existence of equilibrium will further allow for nonincreasing, parametric external, and some cases of internal increasing returns to scale.

TOSHIHIKO HAYASHI, Ph.D. Stanford 1973. Theory of transaction and monetary equilibrium.

Building on a microeconomic analysis of the choice of optimal transactions as a relevant economic problem, this dissertation develops two general equilibrium models of monetary transaction economy. The main conclusion of the study is that it is the desirability of transactions as reflected in the non-Pareto efficiency of the initial allocation of commodities or in the "strong desirability" of transactions that guarantees a positive exchange value for fiat money in equilibrium.

THOMAS K. HOLMSTROM, Ph.D. Oklahoma State 1973. A theoretical analysis of Keynesian labor market assumptions and the macroeconomic adjustment process.

Keynes' labor market assumptions imply a positive

rather than a negative relationship between quantity demanded and the price level. The neo-Keynesian aggregate demand function which is obtained without reference to the labor market is a market equilibrium curve rather than an authentic demand curve. Three new equilibrium curves are presented which help to reveal aspects of the interdependence of the labor market and the other markets in the Keynesian system.

YASUHISA HOSOMATSU, Ph.D. Wayne State 1972. Static and dynamic analysis of Cournot type oligopoly models.

WALTER E. HUFF, Ph.D. Florida State 1972. Economics of the Yugoslav joint venture.

Yugoslavia is a socialistic country seeking ways to improve her economy; one of the more tradition shattering innovations she has introduced is the acceptance of foreign investment in joint ventures with Yugoslav enterprises. This major change in policy was taken in July 1967. Much to the surprise of many Yugoslavs, few investments were forthcoming. The purpose of this dissertation is to explain the initial failure of the Yugoslav Joint Venture Program to attract any substantial amount of foreign investment.

MING-JENG HWANG, Ph.D. Texas A&M 1973. Optimal sizes and shapes of market areas of firms.

This dissertation sets forth a theoretical as well as empirical model on market sizes and shapes, and investigates the sizes and shapes of market areas of firms in the United States. It is contended that small size market areas prevail in the United States. Moreover, it is observed that competitive firms would practice f.o.b. mill pricing in hexagonal shaped markets when market areas are small, and charge discriminatory prices when the areas are large.

EVAN JONES, Ph.D. Michigan State 1973 On the microeconomic theory of optimal capital accumulation.

This thesis explores certain implications of the dynamic neoclassical theory of factor demand; in particular, the problem of economics of scale. An attempt is made to extend the Marshallian concept of long-run average costs to a dynamic framework and to compare the modified equilibrium with those resulting from neoclassical theory. Further, Lerner's accepted interpretation of Keynesian investment theory is argued to suffer from a defect parallel to what in microeconomic theory is a false dichotomy between static and dynamic elements.

SERGEI S. KASAKOW JR., Ph.D. Washington State 1973. A study of managerial incentives in the Soviet Union.

This thesis examines the managerial incentive provisions of 1965 economic reforms for the *USSR*. They are reviewed in relation to Western literature on motivation for work and risk taking. State-chosen success indices (profits and sales) are considered for several firms and industries (Shchekino, Perm's, and Bol'shevichka). All major incentive experiments successfully achieved their planned objectives, but the new approaches were not widely extended, and plant management received minimal cooperation from supervising administrators.

J. THOM KELLY, Ph.D. Texas A&M 1972. A model of price search and allocation of time in the theory of consumer behavior.

A mathematical model is developed in which consumers can find lower commodity prices by investing time in search; prior contributions in price search theory are used in a formal demand-for leisure model. Specifically determined are effects on consumer behavior (leisure, work, and search), of changes in commodity prices, consumer wage rate, search efficiency and income.

MOHAMMED A. KHAN, Ph.D. Yale 1973. Large exchange economics.

MEIR G. KOHN, Ph.D. Massachusetts Institute of Technology 1973. Sequential decision making and the theory of the firm.

PANAYOTIS G. KORLIRAS, Ph.D. Rochester 1973. Essays in disequilibrium macrodynamics.

The theme of these essays is the development of a macroeconomic theory under the nontâtonnement assumption. After a critical review of the related literature, one nonmonetary model and one model of the financial sector are presented. The last essay, integrating both into a general model with variable employment, examines the nature of the adjustment processes and the stability of the short-run equilibrium.

ARTHUR J. KUHN, Ph.D. California (Berkeley) 1972. An application of a system control model to business history: The General Motors Corporation under Alfred P. Sloan, 1920 to 1935.

WAYNE H. LANCASTER, Ph.D. Claremont School 1973. A reconsideration of the Phillips curve for the United States.

JEROME K. LAURENT, Ph.D. Indiana 1973. The development of harbors, waterborn shipping, and commerce at six Wisconsin ports on Lake Michigan through 1910.

This study examines the circumstances which influenced the development of harbors, waterborne shipping, and commerce, at six major Wisconsin ports on Lake Michigan: Kenosha, Racine, Milwaukee, Sheboygan, Manitowoc, and Green Bay. Although initial developments at the several ports (except for Green Bay) took place in the period 1836–65, the era 1866–1910 is especially examined, because the developments of these years determined the pattern of activity of subsequent decades.

DWIGHT R. LEE, Ph.D. California (San Diego) 1972. Dynamic profit maximization under conditions of monopoly.

The basic problem considered is that faced by a monopolist attempting to schedule the production, sales, and inventory holdings of his product over an N-period interval of time in such a way that profits are maximized. The demand schedule faced in each period is assumed known. The solution is restricted by the following constraints: inventory holdings; production and sales must be nonnegative in all periods; and the final inventory must be some specified amount. Only very general restrictions are imposed on cost and demand functions.

EDWARD C. LESNICK, Jr., Ph.D. Notre Dame 1973. A quantitative analysis of the supply and demand for ships: A case study of colonial New York and South Carolina.

The dissertation is a study of the maritime industries of the eighteenth century with the colonies of New York and South Carolina as case examples. Specific objectives are: first, to analyze the supply and demand for ships; second, to organize data and to test certain statistical hypotheses; third, to develop a theoretical model which is capable of being empirically tested; and fourth, to discuss instrumental factors in the economic growth of the colonies.

MICHELLE B. MCALPIN, Ph.D. Wisconsin (Madison) 1973. The impact of railroads on agriculture in India, 1860–1900: A case study of cotton cultivation.

CHARLES P. MCPHERSON, Ph.D. Chicago 1972. Tariff structures and political exchange.

Tariff legislation is viewed as a commodity bought and sold in a market. It is purchased by economic groups in a society (producers or consumers) and is sold by a firm called the political system. A theoretical model of this market is developed and empirical implications are drawn. Cross-sectional and cross-time tests are carried out relating 1954 and 1958 tariff data to industry census data for a number of manufacturing industries. The results are inconclusive.

RICHARD S. MACK, Ph.D. Colorado State 1972. Social rationality and consumer behavior.

ROBERT J. MACKAY, Ph.D. North Carolina 1972. Monetary growth models.

CHARLES F. MANSKI, Ph.D. Massachusetts Institute of Technology 1973. The analysis of qualitative choice.

JAMES R. MARCHAND, Ph.D. Virginia Polytechnic Institute 1973. Two essays on externalities in supply.

This is a dissertation in two distinct but somewhat related parts. The first part of the dissertation analyzes the literature on externalities in supply before World War II. The second part of the dissertation pursues the topic of interindustry externalities in supply through the author's own analysis of this problem.

WARREN T. MATTHEWS, Ph.D. Texas A&M 1972. An empirical test of a theory of spatial pricing.

This dissertation is an empirical test of Greenhut's theory of spatial price structure. Using the concept of cost-imposing economic space between buyer and seller, it is deduced that the firm will practice price discrimination in favor of more distant buyers. The pricing policies of 1200 firms are analyzed to study the impact of costly economic space on spatial price structure, and implications of the theory for *U.S.* antitrust policy are examined.

CRAIG A. MEDLEN, Ph.D. California (Santa Barbara) 1973. Problems of proletarian class consciousness in the writings of Karl Marx.

KENNETH B. MOBERG, Ph.D. Purdue 1972. Individual and social decision processes.

CLAIR E. MORRIS, JR., Ph.D. Wisconsin (Madison) 1973. J. Laurence Laughlin: An economist and his profession.

This is an assessment of Laughlin's contributions to the economics profession with emphasis on his role in the monetary debates of the 1890's and the money reform movement leading to the passage of the Federal Reserve Act of 1913. Attention is also given to his rivalry with Richard T. Ely and the impact which he had on the founding of the "Chicago School."

ANJAN MUKHERJI, Ph.D. Rochester 1973. On the stability of competitive equilibrium.

It is possible to extend and clarify the existing results on the stability of the competitive equilibrium in several directions. The following are treated: a successive tâtonnement where only the good with the largest excess demand has its price raised; extensions of the gross substitute system to include complementarity; stabilizing income effects; variable speeds of adjustment; convergence of prices in a barter framework; dependence of stability results on the choice of numeraire.

JURGEN MÜLLER, Ph.D. Stanford 1972. The impact of information on technical efficiency.

The concept of technical efficiency differences—different levels of output with identical levels of input—is unsatisfactory from a production theoretic point of view. Here we develop a model in which differences in nonconventional inputs and especially information obtained by managers may explain productivity differences between firms. Estimation of the underlying production structure (of a sample of California dairy farms) via a modified nonhomothetic Cobb-Douglas production function show the specific impact of information.

PAUL F. MULLIGAN, JR., Ph.D. Duke 1973. Government functions in an underdeveloped free enterprise economy: The economic system of Liberia.

IWAO NAKATANI, Ph.D. Harvard 1973. Essays in economic theory: Uncertainty, ability, and social welfare.

The thesis consists of three essays. The major theme in all three essays is the analysis of the impact of the existence of uncertainty on the level of individual welfare, where individuals are different with respect to ability, income, and value systems, generally. It is argued that uncertainty is an important source of inequalities of many dimensions. Welfare implications for various classes of individuals which arise from such differences are examined from various aspects.

WAYNE F. PERG, Ph.D. Purdue 1972. Theoretical foundations of macroeconomics.

PIERRE M. PESTIEAU, Ph.D. Yale 1972. Optimal public investment and rate of discount.

GLENN C. PICOU, Ph.D. North Carolina (Chapel Hill) 1972. The neoclassical model of optimal capital demand.

ALAN M. POLINSKY, Ph.D. Massachusetts Institute of Technology 1973. Essays in public sector economics: Central and local.

CLAYNE L. POPE, Ph.D. Chicago 1972. The impact of the antebellum tariff on income distribution.

This thesis examines the impact of the antebellum tariff on income distribution between North, West, and South. A general equilibrium model is constructed which incorporates relevant features of the antebellum economy—intermediate goods, nontraded goods, and factor immobilities. The results of combining plausible structural data with the model indicate that a general increase in tariff rates would redistribute income from cotton planters to manufacturers. However, a tariff increase on cotton textiles reversed this result.

HARVEY J. REED, Ph.D. Purdue 1973. An experimental study of equilibrium in a competitive market.

Equilibrium has been used as a maintained hypothesis in the estimation of supply and demand curves. However, to date little experimental evidence exists for a nontâtanoment process. This thesis attempts to test the zero excess demand hypothesis against an alternative hypothesis which is constructed from a Bayesian learning model for individuals. At least for a small number of periods, equilibrium in the sense that supply equals demand can be rejected in favor of the Bayesian learning model. However, little is known about the consequence of running many more periods.

SCOTT F. RICHARD, D.B.A. Harvard 1972. Optimal life insurance decisions for a rational economic man.

The problems of an individual who is periodically faced with consumption, investment and life insurance decisions are analyzed with a decision theoretic framework. A proper joint utility for consumption streams, legacy, and lifetime is derived under certain behavioral assumptions, and shown to be of product form. For both analytical and numerical formulations of the problem, optimal consumption, investment, and life insurance policies are found by using stochastic dynamic programming to maximize the investor's joint utility.

RICHARD ROMANO, Ph.D. State University of New York (Binghamton) 1973. Dissension within the English classical school: An integration of the thought of William F. Lloyd.

This thesis examines the Oxford lectures (on population, poor-laws, value, rent) of W. F. Lloyd (1794–1852); compares him with Ricardo, McCulloch, James Mill, and dissenters like Bailey, Scrope, Jones; and suggests his lack of influence may have been due in part to a social outlook which often placed him at odds with the process of industrialization and associated him with the views of Paley, Malthus, Jones, and Coleridge, rather than the Ricardians.

JANE L. ROSS, Ph.D. American 1972. A quantitative study of the historical development of England, 1600–1960.

LEONARD G. SAHLING, Ph.D. Michigan 1972. The price equation: An econometric study.

YASUHIRO SAKAI, Ph.D. Rochester 1937. Axiomatic foundations of consumption and production theories.

This study examines four topics: equivalence of the weak and strong axioms of revealed preference without demand continuity assumptions; axiomatic foundations of the indirect utility function, based on the revealed preference approach; an axiomatic approach to input demand theory when a firm produces only one output; substitution and expansion effects in the general case of joint production.

LEE W. SAMUELSON, Ph.D. Harvard 1972. Portfolio theory and the theory of the firm: An investigation of firms' asset management behavior under risk.

Firms are viewed as managing portfolios of assets—both real productive assets and financial assets. A portfolio model of firms' asset management behavior is developed, based on approximate closed form solutions for myopic multiperiod portfolio optimization in the absence of a risk-free asset. Empirical investigation of the model leads to the conclusion that firms' asset management policies are considerably more conservative than would be expected if maximization of (expected) geometric rate of return were their objective.

PHILLIP M. SARVER, Ph.D. Nebraska 1972. Historical influences on the economy of Pueblo, Colorado.

This dissertation (1) examines those critical aspects of growth which have acted to promote or retard the development of the economy; (2) offers an explanation of the present condition of the city; and (3) explores the prospects for the future growth of the community. Although there is now a favorable trend toward greater growth, unless the obstacles which the people themselves have created are overcome, the economic growth will be slow.

MARK A. SATTERTHWAITE, Ph.D. Wisconsin (Madison) 1973. The existence of a strategy proof voting procedure: A topic in social choice theory.

The dissertation considers committees whose task is to select one alternative from a set of three or more alternatives. Committee members cast ballots which are counted by a voting procedure. The voting procedure is strategy proof if it always induces every rational committee member to cast a ballot that reveals his preferences. The main theorem proved states that no strategy proof voting procedure exists that satisfies the Pareto principle and is not dictatorial.

WAYNE J. SHAFER, Ph.D. California (Santa Barbara) 1972. Two essays in mathematical economics.

LAURENCE SHUTE, Ph.D. Columbia 1973. The economic thought of John Maurice Clark: Toward the development of social economics in the United States.

This thesis is a comprehensive study of Clark's works. Pioneering in the institutionalist tradition, his prodigious output spanned a broad range of economics and yielded significant contributions to the formulation of a social, dynamic theory of economic process. Much of his work is now classic: Overhead Costs, Effective (workable) Competition, Multiplier, and Accelerator Analysis, for instance. Nevertheless, there remains a great deal which has not yet found its way into the mainstream of orthodox thought.

AUSTIN SPENCER, Ph.D. Indiana 1973. An examination of relative downward industrial price flexibility, 1870–1921.

This study examines the hypothesis that industries that experienced increased concentration over the period 1870–1921 show a decline in price flexibility, while industries which did not experience increased concentration show no observable reduction in price flexibility. Price data to test this hypothesis were collected for eleven industries that experienced structural change and for two groups of products of industries that remained competitive throughout this period. Comparable price data were collected for New York, Philadelphia, and Cincinnati.

J. R. STANFIELD, Ph.D. Oklahoma 1972. The economic surplus and neo-Marxism: A further revision.

This study is a contribution to the neo-Marxist literature on modern capitalism. The fundamental theme is that neo-Marxism has thus far failed to repudiate sufficiently the competitive law of value. The economic surplus, defined as potential output minus essential consumption, is estimated for the 1929–1970 period, from time-series. These estimates are then used to evaluate certain aspects of neo-Marxism. The major conclusion is that emphasis should be placed upon the composition rather than the level of total output.

IVAN STARKE, Ph.D. Washington 1973. Policy implications of factor and commodity taxes in general equilibrium.

RICHARD J. SWEENEY, Ph.D. Princeton 1972. Firm decision making, and macroeconomic disequilibrium.

In an economy where each of many firms is a monopolist-monopsonist, aggregate business sector decision functions relate *actual* prices and wages and business' *desired* investment, hiring and growth to aggregate demand, the interest rate, expected price and wage inflation, and the aggregate rate of unemployment. Short-run policy recommendations include high aggregate demand to spur growth and hiring, and a low interest rate to mitigate inflation from the cost side.

KATSUAKI TERASAWA, Ph.D. Kansas 1972. Optimal replacement timing and pricing policy of a house.

WILLIAM J. WEEKS, Ph.D. Kansas 1972. An Evans-Roos model for the dynamics of the firm.

The Evans-Roos dynamic models of monopolistic and competitive markets published during the 1920's are the first attempts to apply the calculus of variations to economic theory. The relative obscurity of their work is due to the limited applications of the type of demand curve specified, and to the scanty qualitative information which they extracted from the model. This dissertation is concerned with characterizing the optimal paths of price and quantity under various assumptions about cost and demand conditions. It also develops an alternative dynamic monopoly model for a Wicksteed-type demand curve.

WILLIAM E. WEHRS, Ph.D. Purdue 1972. Incomplete specification and aggregation in a class of econometric inventory investment models.

A class of Accelerator-Buffer Stock models modified for production flexibility is examined with regard to an apparent inconsistency between estimates of two structural parameters. A more complete rational model, consistent with this class, is specified. The expression of the two structural parameters as functions of parameters in the more complete model allows an explicit analysis of the apparent inconsistency. Simulation results tentatively indicate that aggregate coefficient estimates provide little information on micro behavior.

JOHN A. WHIPPEN, JR., Ph.D. Colorado State 1972. Economic and social change.

This study is concerned with the process of trans-

formation of value and institutional parameters as they are reflected in established Western thought. It is proposed that theory is derived from the behavioral norms of society; as these norms evolve theory must also evolve. The key hypothesis of the study is that the positive methodology in economics is conducive to the growth of theory which is incapable of adjusting for value changes in society.

BETTY D. WICKHAM, Ph.D. Rochester 1973. Some operational implications of a linear consumer production model with fewer characteristics than goods.

This paper extends the Lancaster consumer-producer model operationally. It is argued that the existence of linearly related goods implies ruled indifference surfaces and that these relationships can be measured when goods prices cause entire rulings to be efficient. Circumstances leading to this result are examined along with a procedure for estimating the number of characteristics and the objective structure among goods. Finally, a procedure for choosing unique characteristics is proposed.

WILLIAM D. WILLIAMS, Ph.D. Union Graduate School, Antioch 1972. The mathematics of Veblenian economics.

The work of Thorstein Veblen is commonly thought of as fundamentally unsystematic and incapable of mathematical representation. A demonstration is made that a systematic presentation is possible. The essential core of Veblen's thought is the insistence that the fundamental features of economic theory be formulated in terms of change rather than in terms of stationary states. Mathematical representation takes the form of differential equations involving time rates of change of economic variables.

DANIEL P. WILLIAMSON, Ph.D. California (San Diego) 1973. A differential game version of the entry barrier problem.

This dissertation examines the behavior of a monopoly concerned with preventing entry into its market. An entry decision will be based upon the profitability of the postentry duopoly market. The problem of measuring a barrier to entry is that of determining a link between preentry monopoly characteristics and this postentry duopoly profitability. The duopoly problem is analyzed as a dynamic, two person, cooperative game. A key component of such games is the formulation of noncooperative threat strategies which help determine the final arbitrated outcome.

MICHAEL L. WISEMAN, Ph.D. Wisconsin (Madison) 1972. An essay on household investment.

CHRISTINE N. WOLLAN, Ph.D. Illinois (Urbana) 1972, The financial policy of the Soviet State Bank, 1932–1970.

JESS B. YAWITZ, Ph.D. Washington (St. Louis) 1972. Economically relevant wealth: Its specification and role in consumption.

This study investigates the appropriate specification of the wealth variable in the aggregate consumption function. This endeavor is important for domestic stabilization policy, since both the built-in stabilizers and the responsiveness of the economy to policy decisions depend upon how wealth affects consumption. After rejecting the cash balance hypothesis, I measure the net worth contribution of both the federal debt and the money supply. I conclude that the entire debt should be included in household net worth, but only the conventional specification of outside money is relevant for the wealth effect.

JANG HEE YOO, Ph.D. Texas A&M 1972. The general disequilibrium and market adjustment mechanisms in macroeconomic models.

A mathematical model for the general disequilibrium is established recognizing the inflexibility of the price variables. It includes the commodity, bond, and labor market. In particular, Patinkin's spillover effects and Clower's dual decision hypothesis are integrated in the model, and Tucker's disequilibrium model is extensively revised by modifying the concept of the spillover variables. Using this alternative model, the signs applicable to change in the price variables in the various cases of disequilibrium are investigated.

RICHARD M. YOUNG, Ph.D. Pennsylvania 1972. Macro decision making under uncertainty: Some theoretical and simulation results.

Economic Growth and Development; including Economic Planning Theory and Policy, Economic Fluctuations and Forecasting

GEORGE O. ABALU, Ph.D. Iowa State 1973. Optimal investment decisions in perennial crop production in Cameroon: A dynamic programming approach.

Agricultural administrators in developing countries are increasingly involved in making investment decisions within the framework of their national development programs. The dissertation was designed to make some modest contributions to our understanding of (a) the nature of optimal investments in perennial crops, (b) the economic organization of perennial crop production, and (c) the effects of allowing for possible divergences between observed and actual measures of costs and benefits on investment strategies for perennial crop production in Cameroon.

GEORGE T. ABED, Ph.D. California (Berkeley) 1972. Industrialization, employment growth, and economic development (with special reference to the case of Egypt).

ABDEL R. ABDEL-WAHAB, Ph.D. Illinois (Urbana) 1972. Tax policy and problems in a developing economy: A case study of the Sudan.

BASHIR AHMAD, Ph.D. Michigan State 1972. Farm mechanization and agricultural development: A case study of the Pakistan Punjab.

MUSHTAQ AHMAD, Ph.D. Indiana 1972. Production functions in West Pakistan agriculture.

This study focuses on the agricultural crop sector of West Pakistan during the period from 1950–51 to 1960–61. Cobb-Douglas production functions are estimated for both years with cross-sectional data. Land and labor were optimally used in 1950–51 and land and irrigation were almost optimally used in 1960–61. The output elasticity of land has greatly increased whereas those of irrigation and labor decreased. These changes are mainly due to the relative increases in respective quantities. The study suggests that government development of the crop sector should have been based on a greater use of chemical fertilizer and other biological inputs, which increase the per acre yield and substitute for the scarce input of land.

RAIS U. AHMED, Ph.D. Michigan State 1972. Economic analysis of tubewell irrigation in Bangladesh.

CHOONG-YONG AHN, Ph.D. Ohio State 1972. A recursive programming model of regional agricultural development in southern Brazil (1960–1970): An application of farm size decomposition.

M. ATAMAN AKSOY, Ph.D. Boston College 1973. Growth and trade in a two-sector model with nonshiftable capital and endogenous technological change.

GEORGE ALIBARUHO, Ph.D. California (Berkeley) 1973. Effects of marketing board policies on prices, production, and income in Uganda's cotton industry.

GARTH O. ALLEYNE, Ph.D. California (Los Angeles) 1972. A linear programming model of educational planning for Trinidad and Tobago, 1968–1975.

GERHARD ANDERS, Ph.D. Texas A&M 1972. On the economic development of Canada's North West Territories.

Canada spent $80 million annually for NWT development without studying basic economic problems. The dissertation traces how attenuation of property rights among natives modifies income-leisure indifference curves, and yields a negatively sloped labor supply curve. This explains native labor market participation with transfer payments availability. I analyze the influence of a transcendentalist or immanentist worldview on model construction and the determination of policy constraints. Development theories are criticized in light of preceding, and policy recommendations follow.

ROY D. ANDERSEN, Ph.D. Illinois (Urbana) 1972. The financing and allocation of public expenditures in Sierra Leone.

MICHAEL B. ANDERSON, Ph.D. Cornell 1973. A model of the small Chilean firm.

EDWIN K. A. ANG, Ph.D. California (Berkeley) 1972. The use of a social accounting framework of economic growth for long-range projections of the Taiwan economy, 1970 to 1979.

JALAL M. A. BAYATI, Ph.D. Southern California 1973. Appraisal of developmental planning and industrialization in Turkey.

The successes and failures of Turkish economic planning are appraised by analyzing the development of basic industries including mining, iron, steel, and electricity. Each sector is evaluated either by measuring the relative efficiency of total investment in each industry or the returns attributed to the total investment, depending on the availability of data.

ASSEFA BEQUELE, Ph.D. Indiana 1973. Industrialization and labor absorption: Projections for Ethiopia, 1968–1983.

Within the context of the Ethiopian economy, this dissertation examines the interrelations between pace of industrialization and the rates of absorption of surplus labor from agriculture under assumptions concerning population growth, urban expansion, and increases in labor productivity. Required levels of industrialization and values of other macroeconomic variables consistent with exogenously determined rates of growth of employment are defined within the framework of a two-sector projection model based on Harrod-Domar, Mahalanobis, Ichimura, and Chenery models of growth and planning.

RICHARD E. BERNSTEIN, Ph.D. Brown 1973. Factor allocation in tropical agriculture: A theoretical and empirical analysis.

This study analyzes the efficiency of factor utilization on farms of different size and tenure in Brazil and Cuba. Comparisons were based on factor marginal products derived from estimated Cobb-Douglas production functions. The effects of inflation on farm ownership were included in the Brazilian analysis. Data were obtained from agricultural censuses of the two countries. Results demonstrated constant monopsonistic behavior on the part of large landowners.

HANS P. BINSWANGER, Ph.D. North Carolina State 1973 The measurement of biased efficiency gains in United States and Japanese agriculture to test the induced innovation hypothesis.

The hypothesis that biases in efficiency gains arising from technical change and other sources are endogenously determined by economic forces rather than exogenously determined was supported by this analysis. Estimating equations for biases in the many-factor

production case were developed using a transcendental logarithmic cost function in factor augmenting form, making it possible to divide observed share changes into a component due to efficiency gains and a component due to price changes. Components due to efficiency gains were then used to construct indices of biases which were markedly different in the *U.S.* and Japanese experience during the twentieth century.

BARBARA M. BISCO, Ph.D. Cornell 1973. Constraints on economic growth: A case study of the Thai rubber industry.

JAMES E. BLALOCK, Ph.D. Cornell 1973. The economics of the foreign exchange constraint: Theory and a case study of Turkey.

DAVID G. BLANCHARD, Ph.D. Indiana 1973. The impact of external domination on the Liberian mano economy: An analysis of Weber's hypothesis of rationalism.

Max Weber points out that the history of European economic development has been the gradual shift from the traditional (patriarchal and patrimonial) towards the more rational legal domination of economic life. To test Weber's hypothesis this study considers a small group of West African tribes in Liberia who have in the last fifty years begun a shift from a traditional, autonomous, and closed household economy to a peasant and wage-labor economic order operating within a national, political, and bureaucratic domination. A major conclusion is that the transition from traditional to market economy has been speeded up because of outside influences coming primarily from the United States.

WILLIAM A. BOMBERGER, Ph.D. Brown 1973. An error analysis of optimal stabilization policy.

An examination of the performance of various formulations of stabilization policy focusing on the performance of policy in the presence of errors in the policy maker's perception of the economic mechanism. A series of econometric models are selected from the literature and reestimated. Various forms of stabilization policy formulation are derived and their performance simulated in the case in which such a policy is executed by a policy maker who perceives the economic mechanism in the form of one model in a world described by another model or by a model which differs from that of the policy maker in parameter values or lag structure.

JORGE O. BONVICINI, Ph.D. Louisiana State 1972. Cycles in demand-pull inflation: An econometric approach.

SALAH A. BOURJINI, Ph.D. Kansas 1972. Sources of economic growth in developing countries: The case of Tunisia.

RICHARD T. BURCROFF, Ph.D Washington 1972. Contractual choice in Philippine rice farms: Implications for land reform in Asia.

GILES BURGESS, Ph.D. Oregon 1973. A study of modern economic growth using estimates of aggregate production functions for six nations, 1870–1914.

SUTHIPORN CHIRAPANDA, Ph.D. Iowa State 1972. Linear programming models of interregional competition for economic planning of Thai agriculture.

Two major regional planning models for Thai agriculture have been formulated. They are operational but, due to the lack of data, are not applied. The first deals with the short run, utilizing as little uncollected data as possible. It aims to minimize the costs of commodity production, processing, and transportation, subject to land resource constraints and demand requirements. The second model is an extension of the first, involving several resources. It is applicable in the long run. Several policy implications have been made, including water resource development.

KWONG-YUAN CHONG, Ph.D. Michigan State 1973. A simulation policy analysis of the Western Nigerian cocoa industry.

ROGER N. CRAINE, Ph.D. Maryland 1972. Forecasting, adjustment costs and factor demands: Or the emperor's clothes.

This paper constructs a dynamic model of the firm. The demand for four factor inputs—capital, employment, average hours, and the change in inventories—were derived form the theoretical model. The factor demand equations depended on future values of the exogenous variables, which were forecast by the firm. The forecasting procedure gave a distributed lag model in which the weights represented a convolution of the forecasting weights and the weights from the adjustment cost model. Several variants of the model were estimated using a Zellner-Aitken technique to impose restrictions across equations.

ROBERT W. CROWN, Ph.D. Iowa State 1972. A model of income distribution by size-class with application to the results of technical change.

A model is developed that aggregates consistently from the welfare of individuals to that of classes, and a link is made between the derivation of welfare and the participation in factor markets. Two distributions are observable, the functional income and the personal welfare distribution by class. In an application to a district agricultural economy of India, the model is quantified only up to income levels, thus showing the distribution of personal income by class. The impacts of technical change on this income distribution are observed, as are the results of experiments with policy alternatives aimed at attaining equity of income distribution.

JEAN EMILE DENIS, D.B.A. George Washington 1971. The structure of private consumption in selected Latin American countries.

In addition to income level, four other socioeconomic factors are taken into account: relative price structure (for food only), urbanization level, income distribution, and foreign trade. Data on the variables are drawn from published sources, mainly from the UN and ECLA publications. The statistical technique used is a derivation of Venn diagram analysis. It is found that by taking into account the four factors other than income a much better understanding of consumption is achieved than when income level alone is taken into consideration.

FELIX DIAZ-LOPEZ, Ph.D. Southern California 1973. The role of government in industrial development: The case of the Commonwealth of Puerto Rico.

This inquiry examines the role of government in the process of economic progress. Puerto Rico's successful experiment with industrial development provides empirical insights concerning the importance of government as a source of critical inputs to the process of growth. Findings suggest that development policies when properly formulated within a given political and institutional framework, as in the case of the Commonwealth of Puerto Rico, are generally capable of effecting the socioeconomic readjustments necessary for growth.

PAUL H. EARL, Ph.D. Georgetown 1972. An analysis of industrial price behavior.

This dissertation is concerned with the specification, estimation, and simulation of equations to explain price behavior in a representative sample of twenty-two *U.S.* manufacturing industries. The price equations contain cost-based and excess demand pricing hypotheses, derived primarily from competitive and oligopolistic theories. Estimation is by ordinary least squares on data from 1958 to 1969, while simulations are run from 1965 to 1969. Alternative inflation policies with targets consistent with the Economic Stabilization Program are tested in an *ex post* fashion.

LEYLA ECEVIT, Ph.D. Pittsburgh 1972. Structural changes, production functions, and economic development in Turkish manufacturing.

A study of Turkish industrialization during 1950–68. Structural changes across countries and over time are used as a base for comprehensive comparisons with Turkey at the three aggregate sector level and for individual manufacturing industries. Production function analysis is used to identify the relative contribution of various factors of production to changes in output, to differentiate between the performance of the public and private sectors, for preplan and postplan periods.

VASSILIKI ECONOMOPOULOS, Ph.D. Nebraska 1972. A linear programming model and its application to the Greek economy.

This is an investigation of the effects of selected education and manpower policies upon the output of the Greek economy. Using empirical data, the study combines the methods of input-output analysis with those of parametric linear programming analysis. The study reveals the usefulness of such a procedure in policy analysis and implementation.

RICHARD E. EDDY, Ph.D. Illinois (Urbana) 1972. Measuring levels of living.

THOMAS EDENS, Ph.D. Michigan State 1972. Fluctuations in foreign exchange reserves and in the volume of construction: The similarity between industrialized and nonindustrialized countries in 1870–1914 and their divergence in 1955–1968.

This thesis examines the relationship between foreign exchange reserves and the volume of construction in industrialized and nonindustrialized countries. The hypothesis that construction in nonindustrial countries is more sensitive to exchange reserve fluctuations than in industrialized countries is supported by a cross-sectional time-series test. While the development of financial intermediaries to supply funds for housing construction appears to promise employment and better living conditions in the developed countries, this trend may also increase the instability of the construction sector.

AHMED EL-DERSH, Ph.D. Syracuse 1973. Economic sectoral structure, economic policy, and foreign trade: The case of Egypt, 1952–1964.

The subject of this study is the Egyptian economy in the period 1952–64. The purpose of the study is to trace the changes in Egypt's sectoral structure and economic policy and to relate them to changes in its foreign trade sector. The period is divided into three phases: The Free Enterprise (1952–56), The Mixed or Guided Capitalism (1957–60), and The Socialist Central Planning (1961–64).

ULRICH ERNST, Ph.D. Indiana 1973. Human resources in regional economic growth.

The study examines the interactions between human resource formation and economic development in a hypothetical two-region economy. The analysis is based on a series of simulation, with a digital-computer model. This model builds on the microeconomic behavior of three groups of decision makers: students enrolled in the educational system who choose between continuing their education and entering the labor force; individuals employed in one region who choose between staying in that region and migrating to the other; and individuals controlling the regional allocation of investment.

DANIEL C. FALKOWSKI, Ph.D. New York 1972. Nacional Financiera, S.A., de Mexico: A study of a development bank.

As one of the oldest and most successful development banks in the world, Nacional Financiera provides the basis for a case study of one such institution. NAFIN is a unique institution and is as much a product of Mexico's growth experience as it has been a causal factor. Nacional Financiera's outstanding

achievement has been its success in stimulating the development of the domestic financial markets.

MICHAEL FINN, Ph.D. Wisconsin (Madison) 1973. Supervised agricultural credit in Peru: Technique adoption, productivity, and loan delinquency in Plan Costa.

HARTMUT FISCHER, Ph.D. California (Berkeley) 1973. Regional variations of economic development: A case study of Mexico 1940–1970.

ANTONIO FURINO, Ph.D. Houston 1972. Measuring the economic and social factors affecting regional development.

MONDHER GARGOURI, Ph.D. Cornell 1973. Experiments with a dynamic multisector planning model for Tunisia.

NEIL H. GARSTON, Ph.D. Brown 1973. Economic growth and human capital: A theoretical and empirical study.

A two-sector, two-factor, real variable growth model is developed; the factors being physical and human capital. One sector produces consumer goods and physical capital; the other, human capital. With rational investment, the equilibrium growth path is determined by all parameters of the system, including the savings ratio. The model is applied to some available data for the United States since 1929, factor shares in the sectors are derived, and long-run growth rates projected.

JUAN A. GIRAL-BOSCA, Ph.D. Yale 1973. Allocation of resources through planning: The Ecuadorian experience.

LORETTA L. GOOD, Ph.D. Cornell 1972. United States joint ventures and national manufacturing firms in Monterrey, Mexico: Comparative styles of management.

Comparative analysis indicated that growth rates and profitability between 1960 and 1969 were very similar for the two groups. However, the most powerful variables which explained an unusually high degree of within-group variance in performance for the joint venture firms were: distribution of internal financial resources; equity patterns; and investment patterns. However, for the Mexican firms they were: sources of technical information; productivity consciousness; and marketing practices.

ELIAS GUITERREZ-SANCHEZ, Ph.D. Cornell 1973. Factor proportions, technology transmission, and unemployment in Puerto Rico.

CHARLES J. HAULK, Ph.D. Duke 1972. Growth and employment in a labor-surplus economy.

HENRY M. HAYS, JR., Ph.D. Kansas State 1973 The organization of the staple food grain marketing system in Northern Nigeria: A study of efficiency of the rural-urban link.

EMIL O. HERBOLZHEIMER, Ph.D. Michigan State 1972. Cross-section analysis of demand for housing in Venezuela.

The major part of the thesis is designed to determine statistically the most important socioeconomic variables and the value of their parameters with respect to demand for housing in Venezuelan urban areas. Of particular interest is the elasticity of income and of mortgage terms. Tastes and preferences of household heads are analyzed as well as the employment created by the construction of different types of housing units. The data were drawn from three household surveys and mortgage loan applicants.

HENRY HERTZFELD, Ph.D. Temple 1973. Thresholds of regional development in the residentiary industries of the United States.

EKRAM HOSSEIN, Ph.D. Tufts 1973. An econometric analysis of the economic structure and planning decisions in Bangladesh.

WALLACE E. HUFFMAN, Ph.D. Chicago 1972. The contribution of education and extension to differential rates of change.

This econometric research analyzes differential adjustments to disequilibria. The hypothesis that adjustment rates vary positively with decision maker's information levels is tested by fitting a variable adjustment coefficient model between 1959 and 1964 to *U.S.* county aggregate data on nitrogen fertilizer consumption. Measuring information as levels of farmers' education and of agricultural extension activity, the estimated model supports the hypothesis. Increasing education (extension) increases the adjustment coefficient's size; however, increasing education reduces extension's contribution.

THOMAS L. HUTCHESON, Ph.D. Michigan 1973. Incentives for industrialization in Colombia.

PATRICK P. JEANJEAN, Ph.D. California (Berkeley) 1972. Optimal growth with stochastic technology in a multisector economy.

U-JIN JHUN, Ph.D. State University of New York (Albany) 1973. Investment in nonresidential construction in the United States—an econometric study of quarterly time-series—1954–65.

This study develops and estimates investment functions for types of nonresidential construction: industrial, commercial, hospital, educational, religious, and recreational. It makes use of quarterly time-series data for the years 1954–65. The underlying determinants of these types of construction vary considerably, particularly with respect to methods of financing. Unlike residential construction, a separate investment

function is necessary for each component of nonresidential construction.

MUHAMMAD B. M. J. KABBARAH, Ph.D. American 1972. Scope and sequence of industrial development in Syria: An economic analysis.

GOPAL K. KADEKODI, Ph.D. Southern California 1973. Planning for stability and self-reliance: An evaluation of policy approaches for India.

A comparison of two policy models for India suggests that insufficient savings and inelastic demand for investment are the most serious barriers to development. Foreign investment is shown to be usually undesirable because it hurts the balance of payments. India should have the government set guidelines for investment. A two-sector model shows that properly allocated government investment can insure the attainment of a full employment stable growth path in less than fifteen years.

ADIL J. KANAAN, Ph.D. California (Berkeley) 1972. The political economy of development planning: The case of Lebanon.

JOSEPH A. KANE, Ph.D. Temple 1972. The economic significance of development banking in developing countries.

ROBERT J. KESTELL, Ph.D. Wisconsin (Madison) 1973. An analysis of the locational shift of the Brazilian sugar industry.

ASMERON KIDANE, Ph.D. Pennsylvania State 1973. The demand and price structure of selected food products in Ethiopia: An econometric and spectral analysis.

YOUN-SUK KIM, Ph.D. New School 1973. Postwar Japan's foreign trade (1945–1965) and its lessons for developing countries.

Granted that pre-World War II Japan was a formidable industrial country, it still seems justifiable to concede that she could never have made the enormous strides she has taken had she not worked out a very innovative strategy concerning foreign trade. Thus, it is not Japan the industrial nation, but Japan the trade strategist that is the focus of this dissertation. It is the trade strategist that may have much to offer in the way of an example to developing economies.

WERNER LACHMANN, Ph.D. Rutgers 1972. Education and growth: A control theoretic approach.

Most growth models grow at an exogenously determined growth rate. Through the introduction of an education sector which produces technical progress with the help of capital and manpower, the growth rate is internalized. The maximum principle is used to solve the optimization problem. It is shown that the steady state is the optimal state. The disequilibrium analysis shows that a saddlepoint exists and that the optimal control leaks back to the steady state.

OLASUPO O. LADIPO, Ph.D. Michigan State 1973. General systems analysis and simulation approach: A preliminary application to Nigerian fisheries.

JAMES J. LAKSO, Ph.D. Maryland 1973. A technique for choosing between direct foreign investment and domestic investment in production for sale in the domestic market in developing countries.

The dissertation suggests a general, simple decision technique which developing countries can use to evaluate particular kinds of direct foreign investment. The operational basis of the technique is a benefit-cost ratio which compares the benefits and costs of foreign ownership compared to domestic ownership of the same production process. A case study of a foreign investment in South Korea illustrates the applicability of the technique and its sensitivity to the decision variables.

JOHN S. LANE, Ph.D. Stanford 1972. Optimal economic growth and endogenous population change.

In the context of an optimal economic growth model a synthesis of the Ramsey-Meade rules has been presented and the implications examined for an optimal joint savings/population policy. The model used was the Ramsey model with discounting but modified to allow the population growth rate to be determined by per capita income.

JAMES F. LEVINSON, Ph.D. Cornell 1972. An economic analysis of the determinants of malnutrition among young children in rural India.

This research was interdisciplinary in nature examining nutrition and health factors together with economic variables in an attempt to determine more precisely the causes of malnutrition in 6–24 month old children in rural India. Economic status was clearly a most important variable, but improvement in economic status alone would in many cases not improve the nutritional status. The role of educational programs is critically analyzed within the model developed.

WUU-LONG LIN, Ph.D. Stanford 1973. Economic interactions in Taiwan: A study of sectoral flows and linkages.

In Taiwan agricultural development has complemented rapid industrialization as the predominantly agricultural economy has been transformed into a modern diversified economy. This study examines the interaction between agriculture and industry in linkage effects and resource outflow from the agricultural sector during the period 1952–70. Findings are that agriculture can provide momentum for economic growth; the small-scale farm system was able to achieve high agricultural productivity by using labor-intensive and capital-saving technology.

CHUNG-CHI LU, Ph.D. Iowa State 1973. The role of food aid, agricultural development, and capital for-

mation in economic development: A case study of Taiwan.

The major concern of this study was the growth in Taiwan's agricultural output, the formation of capital, the means by which the increased agricultural output was attained, and the role of American food aid in this process. The specific objectives of this study are: to examine economic development of Taiwan's agriculture since World War II and the role of agricultural sector in capital formation; and to appraise and analyze the economic effects of American food aid program in aiding economic development of Taiwan's agriculture.

ANTHONY M. MARINO, Ph.D. California (Santa Barbara) 1972. A two-sector model of growth with asymmetrical productive processes.

JUAN C. MARTINEZ, Ph.D. Iowa State 1973. On the economics of technological change: Induced innovation in Argentine agriculture.

The central objective of this study is to attempt to identify the factors influencing the demand for and supply of agricultural innovations, providing additional explanation for the relative stagnation of the agricultural sector in Argentina for the last decades; in particular, emphasis is given to the Pampean region. The analysis tries then to explain the process of generation of agricultural innovations by public and private institutions and the adoption of the new technologies by the farmer.

AGOGO MAWULI, Ph.D. Illinois (Urbana) 1972. A sensitivity analysis of balance of payments and economic expansion in less developed economies.

CADMAN A. MILLS, Ph.D. Boston College 1973. Development under the minimum time objective; Ghana under Nkrumah.

ESSAM MONTASSER, Ph.D. Princeton 1972. Egypt's pattern of trade and development: A model of import substitution growth.

The study is concerned with Egypt's import substitution growth, the factors underlying its deceleration, and the design of alternative strategies capable of restoring sustained growth. A macro-growth model combining regression and input-output techniques is constructed and used to fulfill these goals. Furthermore, the import substitution growth pattern of structural change is examined through the aid of various empirical structural techniques.

RICHARD W. MOXON, D.B.A. Harvard 1973. Offshore production in the less developed countries by American electronics companies.

This thesis examines the direct investments of American companies in the less developed counties for the manufacture of products and components for sale in the United States. A behavioral model of the offshore production decision process is developed and used to predict the characteristics of products that will be selected for offshore manufacture, as well as the countries that will be chosen as plant sites. The model is tested for the electronics industry.

MOHINDER S. MUDAHAR, Ph.D. Wisconsin (Madison) 1972. Recursive programming models of farm sector with an emphasis on linkages with nonfarm sectors: The Punjab, India.

RANGANATHAN NARAYANAN, Ph.D. Yale 1973. The two-sector model of growth: An essay.

MIAN A. H. NAZ, Ph.D. American 1971. The role of the state in the development of agriculture: A case study of West Pakistan, 1950–1969.

The purpose of this study is to examine Pakistan's agricultural development experience with emphasis on the role of the state. Three hypotheses are investigated: The archaic "state of the arts" may be the reason for low productivity in agriculture; given a stable government, capable of providing the necessary resources and institutional framework, it is possible to change the state of arts and to induce increases in agriculture productivity; and in a developing country technologically induced increases in agricultural productivity can make a significant contribution to overall economic development.

GERALD I. NEHMAN, Ph.D. Ohio State 1973. Small farmer credit use in a depressed community of São Paulo, Brazil.

MOHAMED N. NOOR, Ph.D. Michigan State 1972. The Malaysian small-holder rubber sector: Implications of ethrel stimulation and the new processing technology.

OKONKWO A. NWANI, Ph.D. Florida State 1973. The role of monetary and financial institutions in economic development of Nigeria: 1800–1970.

This study analyzes the role of financial institutions in the development of Nigeria from 1800 to present. Of particular interest was the change in these institutions before and after independence. Prior to independence the structure of banks, their geographic dispersion, and their lending policies were consistent with the support of an enclave economy. After independence there was a marked change in both the structure of banks and their lending policies.

OLABISI OGUNFOWORA, Ph.D. Iowa State 1972. Derived resource demand, product supply and farm policy in the North Central State of Nigeria.

The thesis examined the problems of farm adjustment and increased supply of output in North Central State of Nigeria. The first group of analyses focused on the level of farm income, effect of technology on scale and farm income, and the stability of sole cropping and intercropping systems under different levels of resources and technology. The second group of

analyses consisted of the derivation of resource demand and product supply functions using a combination of parametric programming and multiple regression analyses.

GABRIEL I. O. OJO, Ph.D. Iowa State 1972. An econometric-simulation model of the Nigerian economy.

In this study an econometric model describing the behavior of the Nigerian economy for the period 1951–65 was estimated. The model, which contains fifteen equations, was estimated by two-stage least squares, and was used for simulation and multiplier analysis. The usefulness of this kind of study lies in its capability of broadening the horizon of the policy maker by providing quantitative estimates of the impact of his decisions on the economy.

SUMIYE OKUBO, Ph.D. Tulane 1972. Growth, inflation, and price stability: An empirical analysis.

The purpose of the thesis is to analyze the effect inflation has on economic growth. Although it is often assumed that inflation redistributes income from non-savers to the saving group and thus facilitates growth, high rates of inflation may lead to a less efficient allocation of resources because of uncertainty and other costs associated with inflation and thus may hinder growth. Analyzing a broad spectrum of countries in various stages of development and with various degrees of inflation, we found that mild rates of inflation foster economic growth while rapid rates have a depressing effect on growth.

MARCIA MIN-RON LEE ONG, Ph.D. Ohio State 1972. Changes in farm level savings and consumption in Taiwan 1960–1970.

EMMANUEL N. ONUNKWO, Ph.D. Georgetown 1973. Sensitivity analysis in the evaluation of agricultural projects in Nigeria.

The dissertation has two main objectives. The first an examination of the concept of sensitivity analysis as a theoretical tool for the evaluation, selection, and management of projects. The second an examination of its application to projects in the agricultural sector of a developing country like Nigeria. The dissertation deals with the subject in five chapters, and contains general conclusions on the subject and specific conclusions on the Nigerian agricultural projects subjected to the study.

BOUBAKAR OUATTARA, D.B.A. Harvard 1973. The capital structure problem of development finance institutions: A case study of the Industrial Credit and Investment Corporation of India.

This study of the Industrial Credit and Investment Corporation of India examines three issues concerning development finance institutions: The relationships between capital structure and the cost of capital; lending rates and cost of capital; and the effects of borrowing limitations imposed by foreign lenders on capital structure decisions. The findings were: capital structure and cost of equity are indirectly related; no clear effects of borrowing limitations on capital structure decisions; lending rates are consistent with the cost of capital.

ROOP N. PANDEY, Ph.D. Illinois (Urbana) 1973. Livestock and crop production patterns in Uttar Pradesh, India and projections of alternative livestock development possibilities.

EFSTATHIOS PAPAGEORGIOU, Ph.D. Iowa State 1972. An econometric analysis of the structure of the Greek economy and its development prospects.

The study presents the short-term economic prospects of Greece through the application of an econometric model. First the structure of the Greek economy is examined in an attempt to identify the major constraints to growth. The projections assign considerable weight to the external sector and the sectoral composition of output, reflecting the need for changes in the pattern of output in order to eliminate the consistent and growing deficit in the balance of payments.

IBY A. PEDROSO, Ph.D. Ohio State 1973. Resource accumulation and economics of scale in agriculture—the case of São Paulo, Brazil.

PALANI G. PERIASAMY, Ph.D. Pittsburgh 1972. Economic development in India: The rationale and effects of government intervention since independence.

CLAUDIO POU, Ph.D. Iowa State 1972. Optimal allocation of agricultural resources in the development area of Patacamaya, Bolivia: A linear programming approach.

A linear programming micro-level study of cooperative farms in a setting of transition from traditional to modern forms and methods. Objectives include the determination of optimal land/person ratios, degree and path of mechanization, farm plans, level of resource use and highest achievable income per person. A comparison is attempted between these optimal conditions and the present situation. Policy recommendations are made on the basis of the results. Suggestions for further research are advanced both in agricultural economics and in other agricultural disciplines.

ABUL M. A. RAHIM, Ph.D. California (Los Angeles) 1972. Intersectual equity agricultural taxation and the strategy of economic development: A case study of Pakistan.

GHULAM M. RAHU, Ph.D. Pennsylvania 1972. The role of foreign private investment in the economic development of Pakistan.

SATISH RAICHUR, Ph.D. Pittsburgh 1972. U.S. and USSR economic and military assistance: An empirical analysis.

The study represents a systematic attempt to examine the factors affecting the size distribution of United States and Soviet aid in so far as they can be gleaned from a cross-sectional analysis of aggregate data. The U.S. and Soviet aid flows are related to the economic and political profiles of Third World countries, as well as to their relations with the two superpowers.

JAMES C. RIEDAL, Ph.D. California (Davis) 1972. The industrialization of Hong Kong.

ARCHIBALD R. M. RITTER, Ph.D. Texas (Austin) 1973. The economic development of revolutionary Cuba: Strategy and performance.

Revolutionary Cuba has pursued a changing variety of developmental objectives of an economic nature, the foremost of which have been rapid economic growth, income redistribution, the reduction of unemployment, and the reduction of external "dependence." This thesis evaluates the effectiveness of the changing policies which have been used to achieve these objectives, grouping these policies into three areas: growth strategy, mobilization strategy, and institutional strategy.

CHARLES A. ROBERTSON, Ph.D. Cornell 1973. Economic analysis of ground water irrigation in Nueva Ecija, Philippines.

The objectives of this study include: measurement of the rate of return to investment in irrigation facilities, and identification of factors responsible for variation of rates of return for an area of the Philippines. Alternative measures of rates of return to capital are discussed. Investment in shallow wells is found to be highly profitable, but investment in deep well facilities is in general not attractive.

JAMES A. ROUMASSET, Ph.D. Wisconsin (Madison) 1973. Risk and choice-of-technique for peasant agriculture: The case of Philippine rice farmers.

Several authors have suggested that the Green Revolution has not helped small farmers who are nonadopters to their proximity to subsistence and to the riskiness of modern techniques. This hypothesis is rejected by showing that a risk neutral model predicts actual fertilizer use among a group of Philippine rice farmers better than three alternate models which incorporate aversion toward the risk of "disaster." Several theoretical and methodological innovations may be of independent interest.

D. SAKELLARIOU, Ph.D. Alberta 1972. Industrial linkages: A case study.

ALFRED H. SAULNIERS, Ph.D. Wisconsin (Madison) 1972. Regional allocation of resources: A case study of Peru.

KAMEL S. A. SHOUKRY, Ph.D. Southern California 1973. The economic issues of conflict resolution in the Middle East.

KATAR SINGH, Ph.D. Illinois (Urbana) 1972. The impact of new agricultural technology on farm income distribution in the Aligarh district of Uttar Pradesh, India.

DONALD M. SMITH, Ph.D. Brown 1973. An application of neoclassical growth models to regional growth in the United States.

This study develops a neoclassical growth model to explain growth rates of aggregate income, employment, capital, and net migration rates. The model is tested with aggregate state data, and with sectors (manufacturing, services, agricultural) of state's output. Interregional and factor movements in response to factor price differentials cause wages in relatively low wage sectors to grow at faster rates. However, intersectoral (intraregional) factor movements may reverse this relationship.

GARY H. SMITH, Ph.D. Oregon 1972. Income and nutrition in the Guatemalan Highlands.

WILLIAM G. SNEAD, Ph.D. California (Berkeley) 1972. An economic policy model for the urban people's commune movement in the People's Republic of China, 1958–1961.

JOHN P. STEIN, Ph.D. Chicago 1973. The appreciation of paintings.

Price indices 1946 through 1968 are based on 45,000 auction prices observed in the United States and Britain. The capital asset pricing model is extended to risky durable goods. The appreciation of paintings conforms to the model when the return from nonpecuniary services is 2.1 percent per year. Paintings offer less expected return and less systematic risk than equities. Paintings are "superior" if individuals value viewing services above 2.1 percent, but substantial nonsystematic risk prevents their being "efficient."

HARRY W. STRACHAN, D.B.A. Harvard 1972. The role of the business group in economic development: The case of Nicaragua.

Business groups, long term associations of multifarious firms controlled by multiple families, dominate the private sectors of many developing countries. Their primary function in Nicaragua is financial intermediation; they facilitate both credit and investment opportunities to their members. Lesser services include intragroup business and the insurance of concentrated power. They aid capital accumulation and allocation; they aggressively establish new enterprises; but their concentrated power also brings monopolistic tendencies and other socially undesirable effects.

ESTHER C. SUSS, Ph.D. Pittsburgh 1972. An analysis of the lead-lag relationship between exports, domestic business cycles, and world business cycles: A cross-section study.

This study tries to determine what effect a country's business cycle has on its exports. Export

volume and value were analyzed for eight industrial countries from 1950 to 1970. The major statistical tool used in this study was spectral analysis. It was found that the U.S. and Canada had their export volumes promoted during domestic business expansions, and the U.S. and Austria had their export values promoted during expansions. The other countries had their exports retarded during business expansions.

'D BABATUNDE THOMAS, Ph.D. Indiana 1973. Technology transfer and capital accumulation (a study of the Nigerian manufacturing industries).

The dissertation is focused on the relationship between capital formation and technological progress with the object of determining its impact on value-added per man. The supporting data refers to Nigerian manufacturing industries.

J. PABLO TORREALBA, Ph.D. Michigan State 1973. Improving the organization of fruit and vegetable production-assembly systems in the coffee zone of Colombia: A case study in the La Mesa region.

HIROSHI TSUJII, Ph.D. Illinois (Urbana) 1973. An econometric study of effects of national rice policies and the Green Revolution on national rice economies and international rice trade among less developed and developed countries: With special reference to Thailand, Indonesia, Japan, and the United States.

ALAN T. UDALL, Ph.D. Yale 1973. Migration and employment in Bogota, Colombia.

JAN H. VAN DER VEEN, Ph.D. Cornell 1973. Small industries in India: The case of Gujarat State.

JACOB W. VAN DUIJN, Ph.D. Illinois (Urbana) 1972. An interregional model of economic fluctuations.

SYED H. WAFA, Ph.D. Stanford 1972. Land development strategies in Malaysia: An empirical study.

Two contrasting patterns of land development in Malaysia in operation for a decade provide data for a cost-benefit evaluation. One, the Federal Land Development Authority (*FLDA*), adopted a highly regimented and closely supervised pattern and supplied more capital, technical and managerial skills, and supervisory personnel than the other. The other scheme, the Kelantan State Land Development Authority (*KSLADA*), lacking in capital and skills, relied on settlers' initiative, imagination, and dedication in opening up new land for agricultural settlement. Results of the study indicate that the settler-oriented approach of the *KSLDA* is economically more viable in terms of investment yields and of discounted cash-flow to the government. But the *FLDA* plan provides health clinics, schools, and water supply not possible with the minimal infrastructure of *KSLDA*.

PETER F. WEISEL, Ph.D. Oregon 1972. Changes in traditional agriculture: A study of three towns in northern Liberia.

THOMAS B. WIENS, Ph.D. Harvard 1973. The microeconomics of peasant economy, China: 1920–1940.

The study employs partial equilibrium, microeconomic analysis to explain the interrelationships among economic forces determining the distribution of income in a peasant economy in the early stages of transition to industrialization. Econometric techniques (regression and activity analysis) are applied to production and factor market data pertaining to the Chinese rural economy (1920–40) as a test of the accuracy of theoretical hypotheses.

JEROME M. WOLGIN, Ph.D. Yale 1973. Farmer response to price in smallholder agriculture in Kenya: An expected utility model.

TWATCHAI YONGKITTIKUL, Ph.D. Illinois (Urbana) 1972. Trade and development in Thailand: An analysis of the Industrial Promotion Act using effective rates of protection and linkage criteria.

BASIL ZIMMER, JR., Ph.D. Rutgers 1972. The role of rural to urban migration in the economic development of India.

An analysis is made of the trend and economic effects of rural to urban migration in India, along with the impact of India's Five Year Plans on this phenomenon. Migration is found not to be declining, due in part to shortcomings of rural planning. An alternative strategy is proposed as a method of promoting more balanced regional development and providing a framework in which migration could become a more progressive economic force.

Economic Statistics; including Econometric Methods, Economic and Social Accounting

AHMED H. ABDEL-RAZEK, Ph.D. Kansas 1972. The distributions of least squares estimators in the first-order stochastic difference equations.

WILLIAM E. ALEXANDER, Ph.D. Michigan 1972. An appraisal of the Hedonic approach to the measurement of quality change and a suggested modification of the technique.

AKEEL H. S. AL-SADI, Ph.D. Southern California 1972. Valuation adjustment in input-output tables, economic structure, and economic systems: A comparative empirical analysis.

The Yugoslav and Hungarian economies are compared using price adjustment models that attempt to quantify the degree of distortions in pricing and their consequent effects on the structure of these economies. Distortions in Yugoslavia appeared chiefly in transport and trade, while in Hungary, they were found in

manufacturing, agriculture, and transport. The former apparently resulted from partial price controls and market imperfections, while the latter can be attributed to administrative planning and centralization.

WILLIAM E. AVERA, Ph.D. North Carolina (Chapel Hill) 1972. A geometric mean strategy as a theory of multiperiod portfolio selection.

EDUARD H. BRAU, Ph.D. Duke 1973. A national income, trade-flow, capital-flow model of the United States and the United Kingdom: A simulation study.

WILLIAM R. CRON, Ph.D. Michigan State 1973. The demand for major household appliances: An econometric analysis.

This study attempts to examine alternative approaches to developing demand functions for three major household appliances. Two existing *ad hoc* models are studied to determine their connection with the classical utility maximization framework. Alternative models derived from the utility maximization framework are then developed. In estimating the models a procedure is developed for combining time-series and cross-section data, where the random disturbances are autoregressive, heteroskedastic, and cross-sectionally correlated.

W. ROBERT DALTON, Ph.D. Missouri 1972. Elasticities of substitution in Soviet industry.

In this paper a constant elasticity of substitution (*EOS*) is estimated for nine Soviet industrial sectors for the period 1950–69. Results for all sectors indicate that *EOS*s were significantly less than unity, and that technological change had no downward trend. The *EOS* estimates range from .081 to .671. The analysis graphically illustrates that a slowdown in the growth rate of output is implied by an increasing capital-labor ratio, and a low constant *EOS*.

WILLIAM S. DAWES, Ph.D. Purdue 1972. Some implications of the exact finite sample density functions of *GCL* estimators of structural coefficients in a particular simultaneous equation economic representation.

GEORGE W. GIPE, Ph.D. Oregon 1973. The examination of residuals for the detection of specification errors.

JOHN G. GREENHUT, Ph.D. Texas A&M 1972. A theoretical-empirical study of spatial pricing.

The objective of this dissertation is to formulate a general theory of spatial pricing to include assumptions of variable demands and competition over an economic space. An operational model based on this theory is also formulated to test the theory on observational (assumption) and deductive (result) levels. The impact of a heterogeneous landscape on delivered prices is, in other words, evaluated theoretically and empirically.

CHARLES R. HULTEN, Ph.D. California (Berkeley) 1973. The measurement of total factor productivity in U.S. manufacturing, 1948–1966.

WILLIAM E. JUDNICK, Ph.D. Michigan 1972. A method for modeling multinomial time-series.

GARY P. KUTCHER, Ph.D. Maryland 1972. Agricultural planning at the regional level: A programming model of Mexico's Pacific Northwest.

A programming model of the agricultural activity in Mexico's Pacific Northwest region is developed to assist in policy/planning decisions directly concerning only the region as a substitute for full sector (with regional disaggregation) analysis. The region is large enough to influence national output and input prices and the model is structured to allow for interactions with the rest of the economy through the product and input markets. Simulations of price subsidies, input subsidies, minimum wage changes, and export promotion policies are presented.

OH YUL KWON, Ph.D. McMaster 1972. Optimal fiscal strategies for economic stabilization: An econometric study with illustrative application to Canada.

The thesis is concerned with a particular methodology for applying econometric models to problems of economic policy, with illustrative applications to problems of economic stabilization in Canada during 1967–69. An annual econometric model is estimated, and a policy maker's loss function specified. Government expenditure and personal and corporation income tax rates are chosen as policy instruments. Targets are price change and unemployment. Thirty-six computer optimization experiments are reported, based on various specifications of the loss function.

KYE WOO LEE, Ph.D. Michigan State 1972. An international study of manufacturing production functions: Estimation and implications.

The thesis estimates production functions for three-digit manufacturing industries of the International Standard Industry Classification using combined international cross-sectional and time-series data obtained from the reports of the twelve countries samples. The models used are four generalized production functions discussed recently in the literature. To make a choice among the alternative functional forms, specification error tests are performed. Having estimated the parameters of the production functions, their implications for the international trade patterns are considered.

ANDRE S. LUMBROSO, Ph.D. Pittsburgh 1972. Studies in income distribution.

JOHN MCGEE, Ph.D. Stanford 1973. An econometric analysis of advertising effects.

DANIEL A. MCGOWAN, Ph.D. Pennsylvania State 1973. The measurement of personal wealth in Centre County, Pennsylvania.

GARY M. MULLET, Ph.D. Michigan 1973. Estimators and designs for stationary points on some second- and third-order response surfaces.

KAMBHAMPATI S. R. MURTHY, Ph.D. Western Ontario 1973. Regression by minimizing the sum of absolute errors: A Monte Carlo study.

The sampling properties of the *MSAE* estimator are examined under different assumptions on the nature of probability distribution of the error term, by Monte Carlo experiments. The results confirm the hypothesis that the *MSAE* estimator performs better than the ordinary least squares, in the presence of large errors. The results also indicate the possibility of the unbiasedness and the consistency of this estimator when the errors are symmetrically distributed around zero mean.

BRIAN P. O'CONNOR, Ph.D. Maryland 1973. An income side to an input-output model of the United States.

PHILIP D. OLSON, Ph.D. Oregon 1973. Estimating regression slippage.

DALE J. POIRIER, Ph.D. Wisconsin (Madison) 1973. Applications of spline functions in economics.

CHARLES S. PORTWOOD III, Ph.D. California (Berkeley) 1972. Statistics and their application to social sciences research.

BALDEV RAJ, Ph.D. Western Ontario 1973. The finite sample properties of several estimators of the linear regression models with random coefficients.

The random coefficients regression models are considered appropriate in situations where relevant variables cannot be explicitly included in the regression either because they cannot be measured or other considerations force their omission. This thesis discusses several economic applications of such models. Then, we undertake a Monte Carlo study of small sample and iterative properties of several estimators. Further, we propose new estimators including one which gives guaranteed positive estimates. Several other theoretical extensions are developed.

GOURANGA U. L. RAO, Ph.D. Western Ontario 1973. A Monte Carlo study of small sample properties of selected structural estimators.

The study reports the results of a series of Monte Carlo experiments conducted in order to throw some light on the small sample properties of certain structural estimators. In particular, attention is focussed on the effect produced on the small sample properties of estimators due to the presence of identities and lagged endogenous variables in models.

SUZANNE E. REID, Ph.D. Duke 1973. Corporate simulation: A theoretical approach.

STEVEN C. REIMER, Ph.D. Missouri 1972. Principal components and instrumental variables.

ARTHUR G. REITSCH, Ph.D. Oregon 1973. Statistical tests for detecting the presence of clusters in multivariate data.

HENRY G. RENNIE, Ph.D. Ohio State 1973. Aggregation theory, investment behavior, and rational lag functions.

Three types of measurement error caused by aggregation of investment functions utilizing the rational distributed lag function are presented: Coefficient estimates; lag structures and average lags; estimates of the residual variance. Twenty-seven corporations in six industries for the years 1951–68 were used for each of five investment theories. Aggregation bias at the industry versus firm level is very significant with respect to lag structure and average lag. Bias can also reverse conclusions concerning the relative performance of competiting theories.

MICHEL A. RICHONNIER, Ph.D. California (San Diego) 1972. Experimental design in econometrics: Some extensions.

The context is optimum design for regression analysis. A unifying interpretaion of conventional design criteria is given in terms of concentration ellipsoids. A sufficient condition is given for the optimal number of observations at a design point to be zero. Ordering constraints to be added to the constraint set in a time-series experiment framework are given. The classical orthogonality theorem is extended to time-series experiments, and numerical examples are provided.

KATTAMURI S. SARMA, Ph.D. Pennsylvania 1972. Comparative performance of input-output models with alternative production functions.

VERNON H. SAVAGE, Ph.D. Texas (Austin) 1972. The implications of national defense shifts to a small area: The San Antonio (Texas) case.

Shifting national defense priorities have implications not only to the national economy, but more especially to the regions and industries which have become specialized in this type of activity. This study chooses the San Antonio region and by the use of a Leontief static input-output model, analyzes the implications of such a shift to the region. The analysis is concerned with defense changes and its implications for income, output, employment, and the provision of public goods in the region.

JOSEPH F. SCHULMAN, Ph.D. Michigan 1973. A varying parameter regression study of the tradeoff between inflation and unemployment.

GRAHAM C. SCOTT, Ph.D. Duke 1972. Industrial production and investment in a dynamic, multiregional, interindustry model of the United States.

TERRY G. SEAKS, Ph.D. Duke 1972. An econometric model of the North Pole, 1952–65.

JUDITH E. SELVIDGE, D.B.A. Harvard 1972. Assigning probabilities to rare events.

People's behavior when dealing with the uncertainty surrounding rare events is investigated in order to develop a good procedure for eliciting assessments of the subjective probability of occurrence of such events. Relevant current practices in two settings—specialty insurance underwriting and safety analysis for nuclear power plants—are described and evaluated; results of several experiments in assessing small probabilities are reported; and finally a three-step procedure is recommended for assigning a numerical value to express the uncertainty of rare events.

PHISIT SETTHAWONG, Ph.D. Michigan State 1973. An econometric analysis for Engel curves: Based on M.S.U. Consumer Panel data.

Using M.S.U. Consumer Panel data of 1958, thirteen cross-sectional studies on Engel curves were set out for five composite foods. Income elasticities were estimated based on three alternative mathematical forms of the Engel curves. Cross-sectional and time-series data were pooled. The modified Engel curves were formulated for estimating both income and price elasticities. Some preliminary evidence was presented on the approximate determination of utility functions by means of Engel curves.

M. LYNN SPRUILL, Ph.D. Michigan 1972. Secondary criteria for use with minimum bias estimation of response relationships.

EDWARD VENTO, Ph.D. Texas A&M 1972. Income distribution in a token economy.

WILLIAM E. WECKER, Ph.D. Michigan 1973. Time-series prediction: Theory and practice.

JAMES W. WETZLER, Ph.D. Harvard 1973. Studies on the American distribution of wealth.

This thesis consists of three studies: a survey of the data on the stock of personal wealth in the United States, its size distribution, and time trends in that distribution; measurement of the effective rates of the federal estate and capital gains taxes and the effects of tax loopholes on the rate of return received by the wealthy; and a cross-section study of the wealth-effect on savings using the Survey of Changes in Family Finances.

LAWRENCE C. WOLKEN, Ph.D. Texas A&M 1972. A dynamic price, production, and investment model of a monopoly firm.

Since a static model can only indicate the direction of change of a dynamic system, a dynamic model is developed to trace the firm's optimal path through time. To include the asset variables of the firm in a more direct way, investment is included as one of the control variables, determining the firm's optimal acquisition of capacity over time. A capacity constraint is included, which may restrain the monopolist from taking full advantage of his position.

Monetary and Fiscal Theory, Policy, and Institutions

ABDELRAHMAN T. ABURACHIS, Ph.D. Pittsburgh 1972. Monetary policy, high employment, and external finance.

This paper argues that in order for a central bank target to qualify as the object of central bank control it must meet two specific criteria: the target selected must have a stable and reliable relationship to expenditures; and it must have a stable and reliable relationship to the monetary base. In an aggregate investment equation and in four disaggregated investment equations—two for the business sector and two for the household sector—both the interest rate and the money stock failed to meet the criteria.

DAVID L. ANDERSON, Ph.D. Boston College 1973. A theoretical and empirical analysis of the federal budget process.

Using organization theory and adaptive modeling techniques, a new approach to public budgeting theory is developed. To explain the distribution of funds among defense, human services, physical development, and other budget categories, a simultaneous equation model is constructed and estimated for the postwar period. Structural and inertial forces within budgeting and economic conditions were the most significant explanatory variables; revenues and social factors the least. Decreasing intra-budget funding competition and declining administrative controls over budgeting are found to most hinder effective budget operation.

PANAYOTIS A. APOSTOLIDIS, Ph.D. Arkansas 1973. An investigation of selected operating ratios with emphasis on size of member banks of the Eleventh Federal Reserve District for the years 1960–1969.

Selected ratios—particularly the ratio of earnings to assets—were tested for bank profitability through regression analysis. Priorities in the allocation of resources were identified and marketing policies suggested. Problems of bank performance and scale economies were explored. The findings were compared with similar data from other districts and profit maximization was inferred to be a secondary objective after sales. The study however concluded that profitable coexistence of small and large banks is feasible.

PETER M. ASHTON, Ph.D. Colorado State 1972. Budget allocation in the public sector: A systems analysis of the Colorado Division of Wildlife.

Public finance is an uncomfortable arena in which important decisions tend to be made subjectively without the benefit of systematic economic analysis. A combined *PERT/PPB* cost-effective management system is developed and suggested as a technique for complementing traditional approaches to public budget allocation. The computerized system is evaluated to assess its potential for integrating various critical management functions into a comprehensive decision tool. It is concluded that systems methodology may

significantly enhance public resource allocation decisions.

HUSEYIN S. ATESOGLU, Ph.D. Pittsburgh 1972. Relative stability of monetary velocity and the investment multiplier: Turkey 1952–1967.

LEONARDO J. A. AUERNHEIMER, Ph.D. Chicago 1973. Essays in the theory of inflation.

These are four separate essays on inflation, the first three dealing with aspects of the theory of revenues from the creation of money: the revenue maximizing rate, the welfare costs of inflationary finance, and the stability properties of models where revenue is the exogenous variable. In the fourth essay a model of monetary adjustment is presented which allows for a lagged response of the rate of inflation to changes in the rate of monetary expansion.

EDGARDO BARANDIARAN, Ph.D. Minnesota 1973. The control of money and bank credit in Argentina.

A model of the Argentine monetary sector is developed. It consists of fifteen equations, ten of which explain the stocks of currency, demand, savings and time deposits, legal and excess reserves, ordinary and selective loans, and the interest rates on time deposits and ordinary loans. I present estimations of the structural equations using quarterly data for 1960I–1970IV. The estimated reduced form is then used to analyze the supply of money and bank credit and to evaluate policy instruments. Finally, the implications for the implementation of monetary policy are considered.

THOMAS L. BERTONE, D.P.A. George Washington 1971. Legislative control of executive expenditures in Maryland state government.

The study concludes that the General Assembly of Maryland exercised little control over executive operating expenditures through the budgetary system. The legislature chose not to become deeply involved in budgetary control because such control involves reductions of public services and possible voter retaliation. Broad control over expenditures was obtained, however, by allocating new tax revenues among expenditure categories through the device of funding formulas and special funds.

J. LLOYD BLACKWELL III, Ph.D. Georgia State 1973. A theory of bank structure with applications to problems of differential bank structures, bank pricing policies, bank cost, and bank management.

Models of systems of banks of various structures show that structural changes have welfare effects. The models are also used to compare costs of branch and unit banking, to evaluate branch bank pricing policies, and to develop a model of the optimal branch bank.

MICHAEL D. BORDO, Ph.D. Chicago 1972. The effects of the sources of change in the money supply on the level of economic activity: An historical essay.

The approach of the thesis is to compare different periods in *U.S.* monetary history over the time span 1834–1914 which exhibited different sources of monetary change to see whether the source of monetary change significantly affected the relationship between money and income between these periods as well as within them. In the majority of cases examined, the income effects of the sources of monetary change were found to be insignificant.

JOHN H BOWMAN, Ph.D. Ohio State 1973. Cost and benefit spillouts as factors affecting local taxation for public schools.

The dissertation analyzes factors affecting locally raised school taxes per pupil using linear regression analysis for West Virginia school districts. Ability to levy taxes for local use while exporting part of the burden (generally by taxing business property) was found to have a significant, positive effect on locally raised taxes. Intergovernmental aid extended without local matching or maintenance of effort was found to decrease local taxes; aid imposing such requirements was found to be insignificant.

ROBERT T. BRAY, Ph.D. California (Los Angeles) 1972. Church and welfare property tax-exemption: A study of Roman Catholic Church parish property in Los Angeles County, 1962–1968.

The study presents a model of the multiproduct church-firm based primarily on a Roman Catholic Church parish. Then, using the tax and assessment data of Los Angeles County, the study presents data for the measurable costs and benefits of the tax exempt status of Roman Catholic Church parish property in the county. And, the validity of church property assessments in the county is investigated. The study also includes a discussion of the immeasurable costs and benefits of church property tax exemption.

FREDERICK S. BREIMYER, Ph.D. Northwestern 1972. The money stock and the impact of high-powered money: An aggregated and regional investigation from 1879–1897.

The development of monetary institutions and relationships governing monetary movements were examined for the United States from 1879–1897. It was found that high-powered money explains less variation in money than explained by gold flows. Similar findings occurred using regional deposit breakdowns for national banks. The net deposit definition requiring reserves was discovered for each year, with the determination made that the official net deposit and required reserve data were incorrectly calculated and biased downward.

WYNN V. BUSSMANN, Ph.D. Wisconsin (Madison) 1972. Estimation of household corporate stock portfolios and a model of rate of return from a sample of Wisconsin income tax returns (1946–1964).

HARRY F. CAMPBELL, Ph.D. Queen's 1973. A benefit-cost rule for an additional public project in Canada.

ALEXANDER CASSUTO, Ph.D. California (Los Angeles) 1973. The stability of velocity and the effectiveness of monetary policy in developed and developing countries.

CHARLES D. CATHCART, Ph.D. Virginia 1973. Essays in monetary dynamics.

Three essays present new results on: The time path of the rate of monetary expansion and the stability of the monetary system in hyperinflation when the revenue from inflation is constrained; the excess burden and efficiency of inflationary finance when transitional dynamic effects and a positive growth rate of real income are introduced; and the stability of the monetary system when the rate of monetary expansion is fixed at a low level.

CLAUDIO CONTADOR, Ph.D. Chicago 1973. Money, inflation and the stock market: The Brazilian case.

GEOFFREY R. CONWAY, D.B.A. Harvard 1973. The supply of, and demand for, Canadian equities.

This study provides new statistical information on the market for Canadian stocks—size, ownership, new issues, future demands, etc. Some traditional views, i.e., that Canadians are conservative investors, are shown to be inaccurate. A major foreign investment factor illustrated is that almost 80 percent of the largest Canadian private companies are foreign controlled. The major conclusion is that Canadian institutional demand for stocks will exceed the Canadian supply unless many private companies go public.

TIMOTHY Q. COOK, Ph.D. Purdue 1973. The structure of long-term interest rates.

GILBERT L. COURSE, Ph.D. Purdue 1972. An investigation of the theory of money.

JAMES R. CROTTY, Ph.D. Carnegie-Mellon 1973. The objective function of the state: A study of the political economy of monetary and fiscal policy.

ADROALDO M. DA SILVA, Ph.D. Chicago 1972. Inflation and the demand for money: A four country model.

ROBERT T. DEACON, Ph.D. Washington 1972. The demand for services of local governments: Theory and evidence.

JAMES W. DEAN, Ph.D. Harvard 1973. Aspects of instrument/target links in Canadian monetary policy.

This dissertation examines certain links between instruments and targets of Canadian monetary policy. Reserves-to-money-supply links are analyzed in a money-multiplier context, instruments-to-reserves links are approached via estimates of the effectiveness of defensive policy, and feedbacks of targets to instruments are treated in a reaction-function context. Finally, the link between the secondary reserves instrument and its targets is analyzed using a model in which banks's utility depends upon income and liquidity.

DONALD M. DEPAMPHILIS, Ph.D. Harvard 1973. A microeconomic econometric analysis of the short-term commercial bank adjustment process.

The principal intent of this thesis is to construct a microeconomic model approximating commercial bank adjustment behavior in a period as short as a week. A set of behavioral relationships describing such adjustments is derived and estimated for a sample of seven large commercial banks. The model focuses on the activities of the money position manager in selecting the optimal "adjustment" portfolio. A specific emphasis is placed on the role of nondeposits sources of funds in this process.

J. KURT DEW, Ph.D. Texas A&M 1972. Generalized utility and cardinality price theoretic models of costs of trade.

LARRY L. DILDINE, Ph.D. Michigan 1972. The relation of local public service benefits to property values: Evidence from referenda.

WALTER DOLDE, Ph.D. Yale 1973. Capital markets and the relevant horizon for consumption planning.

STEPHEN M. DUCKWORTH, D.B.A. Harvard 1973. Post-CD liquidity management: A study of New England regional commercial banks: 1960–1972.

The dissertation investigates the effects of liability management on liquidity management practices of selected regional banks. Both determinants and financing of secular cyclical and short-term liquidity needs are analyzed and explained through statistical and interview data. The 1972 policies for discretionary bank assets and liabilities are developed with special attention to the factors effecting levels of temporary investments and the combined CD/net Federal Funds position.

DAVID B. EASTWOOD, Ph.D. Tufts 1972. Consumer credit and consumer durable expenditure.

MOHAMED M. EL-SHESHAI, Ph.D. Indiana 1973. Chance-constrained model for liquidity management in commercial banks.

This thesis studies factors affecting a bank's liquidity position. It describes how a commercial banker may act as a decision maker in a world of uncertainty. A key hypothesis is that the decision maker maximizes expected return subject to constraints dictated by a "tolerable" level of risk. The study develops a model designed to improve efficiency in managing the bank's liquid assets, compared to the usual "rule-of-thumb" one encounters. Efficiency is judged on the basis of cost and ability to meet Federal Reserve requirements.

DAWN E. ELVIS, Ph.D. Michigan 1972. The monetary sector of Jamaica.

DAVID ERICKSON, Ph.D. Florida State 1972. Analysis of the pricing and profitability of put and call options.

The study investigates the economic nature of the put and call option market, the pricing of options, and the profitability of alternative investment strategies. It was found using an option data set of 7416 actual transactions including both "bot" and "sold" premiums that option price can be explained by a log-linear function of duration, volatility, and price. Purchasing straddle options was found a viable strategy which can be expected to generate superior returns.

WILLIAM H. FOELLER, Ph.D. Iowa State 1972. A method for evaluating the impact and incidence of state and local taxes with applications to incidence and educational grant-in-aid programs in Iowa.

This dissertation uses the concept of legislative intent to assign the impact of state and local taxes of a state's residents and nonresidents. Then by various theoretically supported shifting assumptions, the incidence of state and local taxes is determined. An application of the procedure yields estimates of tax burdens on Iowans and effective rate patterns among income groups. Further application of the study method yields differential tax incidence of various state educational grant-in-aid programs.

BERNARD H. FRERICHS, Ph.D. Washington State 1973. A proposed public good pricing system for metropolitan areas.

This thesis develops a system of user charges for urban public services to make the property tax reflect the demand for public services. The goal of this tax, which is generated by a multi-dimensional tax base, is economic efficiency. Multiple regression analysis is used to formulate the relationship between public spending and characteristic variables. Data derived from municipalities within the Seattle *SMSA* are used to test the feasibility of constructing the relationships. This system of taxation appears to be feasible.

CARL M. GAMBS, Ph.D. Yale 1972. The economics of an automated payments system.

JAMES F. GATTI, Ph.D. Cornell 1972. Determinants of the percentage rate of growth of the United Kingdom money supply: 1950–1968.

RICHARD F. GLEISNER, Ph.D. Georgetown 1972. Short-run estimations of changes in the money multiplier and the money stock, 1957–1970.

First differences in the money supply are subdivided into the contributions due to changes in the monetary base and changes in the money multiplier. First differences in the money multiplier, in turn, are subdivided into the contributions of its determinants. Changes in the determinants are then estimated by least squares regression analysis. The R^2 and standard error of estimate of the estimations of changes in the money stock are significantly improved over estimations based on the assumption that the money multiplier retains its value of the previous quarter.

ESEN Z. GURTUNCA, Ph.D. Purdue 1973. A financial growth model.

In this thesis a Keynes-Wicksell monetary growth model with disequilibrium in security and commodity markets is constructed. The money market is not explicitly considered, but it is explained by reference to Walras' Law. After having constructed the model, neutrality of money and the effect of a stabilization policy on balanced growth equilibrium are examined.

JOHN P. HANSEN, Ph.D. Rice 1973. Should tax rates really be considered endogenous?

WILLIAM B. HARRISON, Ph.D. Maryland 1972. Institutional and cash flow variability effects on commercial bank spot accounts.

This thesis examines certain liquid assets and borrowing ('spot") accounts of 350 Fifth District banks. Moreover the relationships between these accounts and the particular sources of deposit and loan variability are studied. The institutional characteristics of banks serve as proxies for nonmeasurable cost differences from bank to bank, and these, too, are found to affect the relative levels of spot accounts. Discriminant analysis helps identify characteristics which distinguish borrowing from nonborrowing banks.

ARLINE A. HOEL, Ph.D. California (Los Angeles) 1973. The effect of reserve requirement changes on bank wealth 1913 to 1960.

Under certain conditions economic theory implies a negative relation between reserve requirement changes and bank equity changes. Bank stock prices, used as a proxy for equity, were collected for all reserve requirement changes between 1913 and 1960. The data were subject to the Chi-Square sign test, the t-test of means, and linear multiple regression analysis. The test results do not provide unambiguous evidence to support or reject the implied negative relation.

W. SLATER HOLLIS, Ph.D. Mississippi 1972. The economic costs of alcoholic beverages to governments and taxpayers: A study in public finance.

This interdisciplinary study hypothesized that the economic costs of alcoholic beverages to state and local governments are greater than the economic revenues received. Subordinate hypotheses were that alcohol is "the" or "a" primary cause of crime, accidents, disease, divorce, underemployment, and pollution and that they involve economic costs to governments. It was concluded upon testing that the hypotheses were valid and that the costs to governments are greater than the revenues by four to one.

PETER W. HOWITT, Ph.D. Northwestern 1973. Studies in the theory of monetary dynamics.

Three studies in disequilibrium monetary dynamics. Walras' tâtonnement theory is criticized as a conceptual framework for analyzing monetary price dynam-

ics. The stability of the long-run quantity theory is examined using a shopkeeper system that can serve as a general non-Walrasian framework for monetary dynamics. Keynesian short-run behavior is shown to occur out of equilibrium even if the long-run quantity theory is valid. This approach provides a nonmonetarist explanation of the observed correlation between money and income.

DANNIS J. JACOBE, Ph.D. Virginia Polytechnic Institute 1973. The National Dividend Plan: A public choice analysis.

A proposal called the National Dividend Plan has received considerable attention. The plan involves a proposed constitutional amendment which would limit the corporate income tax rate; make dividends nontaxable; and redistribute corporate income revenue to persons voting in the last preceding national election on a per capita basis. This study represents an investigation into the implications of the *NDP* by considering each of its most significant provisions, thus aiding the reader in evaluating the plan and answering for himself the question of its overall merits.

KAREN H. JOHNSON, Ph.D Massachusetts Institute of Technology 1973. Essays in monetary theory and policy.

RAYMOND M. JOHNSON, Ph.D. Oklahoma State 1973. The empirical question of money illusion in the United States: Its implications for a Patinkin type model.

The present study considers the implications which the presence of money illusion has for Patinkin's model, and estimates the degree of money illusion present in the economy. This involves the specification and estimation of a consumption function which uses the Kane-Klevorick price parameterization technique to estimate the degree of money illusion. The model is based on Patinkin's specification and includes permanent income, the interest rate, and real balances as independent variables.

JOHN F. JOHNSTON, Ph.D. Duke 1972. Toward a positive theory of redistributional public expenditures.

DAVID H. JONES, Ph.D. Texas (Austin) 1972. The transmission of the effects of an open market operation.

Ratios of their "quick" funds net to their total deposits-net—credit tightness ratios—were calculated for the weekly reporting banks in each Federal Reserve District and for those in New York City for the period July 1966 through December 1972. The credit tightness ratios of the weekly reporting banks in each Federal Reserve district is then expressed as some function of the credit tightness ratio of the N.Y.C. weekly reporting banks. Using the polynomial distributed lag model, final estimating equations for the credit tightness ratios of the twelve Federal Reserve Districts were obtained.

MARTIN E. JUDD, JR., Ph.D. Oklahoma State 1973. The demand for money by state and local governments.

In this study, widely accepted aggregate demand for money hypotheses are tested for the state and local government sector, the states of Washington and Wisconsin, and Northampton County, Pennsylvania. The periods of the study are: state and local government sector, 1957–68; state of Washington, 1957–68; state of Wisconsin, 1960–66; and Northampton County, 1964–68. The testable hypotheses are derived from the Keynesian motives approach to the demand for money.

HERBERT M. KAUFMAN, Ph.D. Pennsylvania State 1972. Price expectations and the demand for money: The American experience of creeping inflation, 1967–1970.

SAM R. KEELEY, Ph.D. Oklahoma State 1972. The existence of a long-run reaction function of the monetary authorities.

Implicit in the concept of a reaction function is the fact that it must apply on a temporally consistent basis. Since stability per se has received essentially superficial treatment, it was the purpose of this study to investigate whether the evidence supported the existence of a stable reaction function of the monetary authorities for the period 1951III to 1969IV. The stability tests employed were moving regression analysis to which confidence intervals were applied, the Chow test, and the generalized dummy variable test. Temporally stable reactions to employment and price stability objectives were found.

JAMES M. KENNEY, Ph.D. Stanford 1972. Credit availability and consumer investment in durable goods: The case of automobiles.

An aggregate structural model of the automobile loan market is formulated and tested (using quarterly time-series data for 1957–70) in order to assess the quantitative impact of changes in monetary policy on automobile expenditure. The lagged adjustment of the automobile loan rate is posited and is substantiated by the data. The resultant need for nonprice rationing in the automobile credit market is found to magnify the short-run impact of tight monetary conditions.

CHANDRA P. KHETAN, Ph.D. McMaster 1973. Econometric model of the Indian monetary sector.

A quarterly econometric model of the monetary sector of India is constructed. The model estimates using quarterly time-series for national income and its components are developed. Eight behavioral equations are estimated for the demand and supply of six principal financial assets. Seasonal variations in some monetary aggregates are dependent on seasonal variations in agricultural income and output. The price level appears sensitive to changes in the high-powered money supply.

DONALD W. KIEFER, Ph.D. Purdue 1972. Economies of scale in the provision of urban governmental services: An examination of city-county consolidations.

This study develops and illustrates a technique for examining governmental reorganizations to determine whether scale economies are achieved. The technique is illustrated by application to two public services in two city-county consolidations. The public services analyzed are government administration and primary and secondary education; the consolidations are in Virginia Beach, Virginia, and Nashville, Tennessee.

IN KIE KIM, Ph.D. Maryland 1972. The sources and changes in the stock of money in the case of Korea 1954 to 1969.

GREGORY C. KROHM, Ph.D. Virginia Polytechnic Institute 1973. An economic analysis of municipal annexation policy: Approached from the theory of fiscal clubs.

Annexation is one of the major means by which the service areas for local public goods are altered. This dissertation examines the variation in state statutes dealing with annexation. It attempts to interpret the legal criteria for annexation in the light of academic theory, especially recent theories dealing with the optimal organization of local "fiscal clubs." Although all states are surveyed, special emphasis is placed upon the particularly interesting case law which has evolved in Virginia.

FINN E. KYDLAND, Ph.D. Carnegie-Mellon 1973. Decentralized macroeconomic planning.

JULIAN E. LAGRAND, Ph.D. Pennsylvania 1972. An investigation of the utility function implicity in government taxation and expenditure decisions.

DANIEL J. LARKINS, Ph.D. Georgetown 1973. Defensive and dynamic lags in open market operations.

A reduced form model is used to estimate the lags between *OMO* and changes in the money stock during the 1960's. The results suggest that, contrary to the findings of other investigators, the lags are not so long as to vitiate discretionary monetary policy.

JOHN E. MCCAIN, Ph.D. Oklahoma State 1973. Expenditure response of local governments to external grants-in-aid.

This study examined the question of how local government expenditure levels on major functions changed as a consequence of receiving external grants-in-aid. An answer was sought as to whether aid recipients substitute outside funds for local funds, or whether the aid acts as a stimulant to expand local spending from locally raised resources. Hypothetical responses were tested empirically by the use of statistical linear regression analysis to estimate the actual expenditure response, and the extent to which this response conformed or deviated from the a priori theoretical model.

MARVIN MCCLAY, Ph.D. Oregon 1973. Comparative investment performance of bank pooled trust accounts and mutual funds.

J. HUSTON MCCULLOCH, Ph.D. Chicago 1973. An estimate of the liquidity premium.

The liquidity premium is estimated using the assumption of unbiased forecasting. Although the expectations hypothesis is far from applicable for forward interest rates of short duration, it cannot be shown to be wrong for forward rates of long duration. When a form that increases at a decreasing rate is imposed on the liquidity premium as a function of maturity, the premium becomes virtually negligible for forward loans of more than a few years duration.

MELVILLE L. MCMILLAN, Ph.D. Cornell 1973. Jurisdictions, grants, and public good supply in a fiscal federalism.

EDWARD MALCA, Ph.D. City University of New York 1973. Performance and characteristics of bank-administered commingled equity funds for employee benefit plans, 1962–1970.

This study investigates the performance and characteristics of thirty-seven commingled pension funds during the period 1962 through 1970. Employing methodology similar to the recent SEC study, it was found that, adjusted for risk (the mean *beta* was .96) average portfolio return for these funds was a 1.6 percent per annum worse than pure random selection would have produced. Of the thirty-seven funds, thirty did worse than totally "unmanaged" portfolios. It was also found that portfolio turnover, bank size, fund size and size growth rate were irrelevant to performance. Also there was no consistency in year-to-year performance by any of the thirty-seven institutions.

CLAYTON G. MAMMEL, Ph.D. Kentucky 1973. The short-term fluctuations of regional housing and mortgage markets.

This dissertation is concerned with the problem of explaining regional variations in the short-housing cycle. The focus is on the financial aspects of these fluctuations as they are manifested in regions.

JAMES B. MARSH, Ph.D. Chicago 1972. A theory and estimation of gold supply: The case of South Africa, 1956–1966.

The study integrates the theories of multiple outputs, natural resource exhaustion, and gold supply into a general theoretical model of mining production. The model is then tested against data from the South African gold-uranium industry, and short-run gold supply elasticities are computed.

GEORGE R. MEADOWS, Ph.D. Washington (St. Louis) 1972. Residential property values and local public budgets: An empirical study of the capitalization of local property tax and expenditure differentials.

This study investigates the capitalization of inter-

community differentials in property taxes and local public services into the value of residential dwellings. A household location model illustrates the expected theoretical relationship. The empirical model proceeds from the concept of housing as a composite commodity defined over dwelling and community characteristics. The empirical results are generally supportive of the capitalization hypothesis. The results indicate that certain property tax reforms may result in important changes in property values.

LAWRENCE J. MURPHY, Ph.D. McMaster 1973. An analysis of the demand for Government of Canada marketable debt.

This thesis provides a systematic analysis of the portfolio behavior of the principal holders of Government of Canada securities. Nongovernment participants in the market for these securities are chartered banks, nonbank financial institutions and nonfinancial corporations. A theoretical model of portfolio behavior is specified. Changes in the portfolio structure of these institutions over time are examined. Econometric techniques are used to evaluate the principal determinants of the demand for Government of Canada securities by each institution.

WILFRIED R. MUTH, Ph.D. Southern Methodist 1973. An econometric model to evaluate monetary and fiscal policies in Germany.

This dissertation presents an econometric model for the evaluation of monetary and fiscal policies in West Germany. For the estimation, quarterly data are used covering the period 1957–70.

FOREST E. MYERS, Ph.D. Kansas State 1972. Differential industry impact of monetary and fiscal policies: An interindustry model.

FERNANDO E. NARANJO, Ph.D. Pennsylvania 1973. A macroeconometric model of fiscal policy for Costa Rica.

JOHN A. NERI, Ph.D. Maryland 1973. Income expectations and the accumulation of real cash balances.

Short- and long-run income expectational elements are combined and applied to money demand. Expected income hypothesis is a generalization of adaptive expectation scheme in that adaptive, regressive, and extrapolative forces are combined. The objective is to determine whether holdings of cash balances are adjusted to expected short- or long-run income position. Distributed lags on current and past values of income are estimated via polynomial interpolation. I conclude that holdings of real balances are adjusted to short-run rather than long-run income position.

DOUGLAS A. NETTLETON, JR., Ph.D. Oklahoma State 1972. The impact of deposit variability on average labor cost in commercial banking.

This dissertation seeks to ascertain whether deposit variability is a determinant of average labor cost in commercial banking. Cost and deposit level data from the Federal Reserve reports of a sample of forty banks were used. So as to distinguish the impact of deposit variability on average labor cost from the impact of bank size on average labor cost, banks in the sample are divided into five size categories. Cost functions are then estimated. Average labor cost was positively related to deposit variability.

ELDON J. NOSARI, Ph.D. Kentucky 1972. National bank entry.

This study argues that fixed entry standards are a more desirable method of controlling rates of entry into banking. If fixed entry standards are to be utilized, knowledge of the determinants of desired entry into banking is required. This study is an initial attempt to investigate such determinants through the development of a national bank entry function. The main conclusion is that the expected rate of return in banking derived from historical data is statistically significant in explaining the variability of desired national bank entry.

LOUIS E. NOYD, Ph.D. Oregon 1973. An empirical test of government's demand for money: The local government case.

TOM OBERHOFER, Ph.D. Rutgers 1973. The redistributive effect of the federal income tax: A time-series and cross-section study.

The redistributive effect of major federal personal income tax components is estimated from time-series and cross-section samples. The Gini coefficient measures income concentration. Exemptions, combined exemptions and deductions and tax rates contribute to the progressivity of the tax while the inclusion of deductions decreases the progressivity of the tax. The redistributive effect of tax components varies as sample mean income varies.

BRUNO A. OUDET, Ph.D. Indiana 1972. An examination of returns on financial assets during periods of inflation.

Hypotheses on the variation of returns of financial assets in periods of inflation are formulated from the existing theories and tested empirically. The justifications of the proposition that stocks are a good hedge in periods of inflation are reexamined, and it is shown that various time patterns of variations of stock prices are in fact possible. The analysis reveals that, whereas in the long run stocks do provide a perfect hedge against inflation, falling stock prices can be quite consistent with the theory in the short run.

SEUNG OH PARK, Ph.D. Kentucky 1972. The tax exporting adjustment in interstate comparison of tax burdens.

Business and personal shares of state-local tax revenue are estimated for fifty-one states for 1967. Because business tax exporting can be influenced, but importing cannot, states with lesser emphasis on busi-

ness taxes and/or lower business values per capita have family tax burdens that are higher than is suggested by the taxes-per-capita measure.

JOSEPH E. PLUTA, Ph.D. Texas (Austin) 1972. Growth and patterns in U.S. government expenditures, 1956–1970.

This study analyzes the growth and changing patterns in *U.S.* public expenditures between 1956–70. It updates past budget studies and combines the scattered literature of the past fifteen years on aspects of the public budget. The growth in outlays is examined to determine the relative shift in priorities to domestic areas. To the extent that relative changes in the budget have proceeded in the direction of education, income maintenance, housing, public health and have been the result of intense federal legislative action, this change in priorities is termed a Social Revolution.

ELLIOT PONCHICK, Ph.D. Southern California 1973. The effects of monetary and fiscal policy on the distribution of wealth in America.

This research is an inquiry into the wealth distributional effects of government policy that operate indirectly through interest rates and the price level. To find how wealth is distributed an individual's portfolio selection process is explored. This selection process is a function of the amount and kinds of information concerning market activities he possesses. It is shown that the larger the portfolio the greater the incentive to acquire information. Optimal purchasing times are earlier and returns greater, the larger the portfolio.

MARY E. RIEDER, Ph.D. Iowa State 1973. A simultaneous equation model to determine the rate of inflation.

This study concerns itself with the direct effect of changes in the money supply on wages and prices in the context of a simultaneous equation system. To do this an eight-equation quarterly macromodel of the United States was formulated and estimated from 1952 to 1969. The model was taken into its derived reduced form and the matrix of reduced form coefficients computed from which several observations about the economy were made.

EDMUND J. ROLLO, Ph.D. West Virginia 1972. The financing of highways in West Virginia with suggestions for reform.

JAYANTA ROY, Ph.D. California (Berkeley) 1972. A model highlighting fiscal and monetary policy effects in India 1950–1965.

FRANK S. RUSSEK, JR., Ph.D. Maryland 1972. The Tax Reform Act of 1969: A study of its relative impact on housing investment.

This study estimates the differential effect of various tax reforms on investment in housing and other durable assets. Micro-simulation models are used to calculate the tax-induced changes in the present value of apartment buildings, office buildings, and producers' equipment. Macro-simulation experiments are performed with a variant of the FRB-MIT econometric model to determine how different tax provisions alter investment in housing, producers' structures, and equipment when overall investment is held constant.

MICHAEL SARRIS, Ph.D. Wayne State 1972. An analysis of the findings for monetary and fiscal policy in reduced form macroeconomic models.

ROGER W. SCHMENNER, Ph.D. Yale 1973. City taxes and industry location.

JOHN B SHOVEN, Ph.D. Yale 1973. General equilibrium with taxes: Existence, computation, and a capital income taxation application.

WILLIAM G. SMILEY, Ph.D. Iowa 1973. The evolution and structure of the national banking system, 1870–1913.

The period of 1870 through 1913 was one of rapid economic development and industrial transformation in the history of the United Utates. The role of the national banking system in facilitating this development and transformation has long been the subject of debate. It was often suggested that the national banks were relatively ineffective in transferring financial capital from the more industrialized regions—the New England, Middle Atlantic, and later, East North Central regions—to the more rapidly developing southern and western regions. The present study is concerned with developing a basis for more accurate evaluation.

SUNG WON SON, Ph.D. Pittsburgh 1972. Certificates of deposits in liability management.

The first two chapters are devoted to the discussion of the development of various money market instruments and their contributions to commercial bank liability management. The core of the dissertation is the third chapter which develops the theory of liability portfolio selection using the mean-variance approach to Tobin-Markowitz. Chapter four attempts to obtain empirical support for the theory developed earlier. Finally, the entire study is summarized in chapter five.

LARRY V. ST. LOUIS, Ph.D. Oregon 1972. Price expectations and the term structure of interest rates.

HOWARD K. STOKES, JR., Ph.D. Georgetown 1973. Portfolio theory: The estimation of indifference curves for government securities.

The dissertation attempts to estimate a risk return preference function in the Tobin-Markowitz tradition. Data for six maturity classifications and seven financial intermediaries were taken from the U.S. Treasury Survey of Ownership of Government Securities. A *CES* utility function was used in the estimation. The results indicate a stable, but yet concave, preference function. An extension is then presented to the analysis of the term structure of interest rates.

GERALD W. STONE, JR., Ph.D. Rice 1973. The fiscal

feasibility of land value taxation in urban metropolitan communities.

RAINER STUPER, Ph.D. Florida State 1973. An empirical analysis of the debate over rules versus discretion with special reference to the monetary management of the German Bundesbank from 1958 to 1970.

An attempt to evaluate performance of strategies propounded for monetary policy which use the money supply as indicator. Strategies considered are "fixed" rules, which confine the growth of the money stock to some constant rate; "flexible" rules, which employ formulas to establish required change in money stock and discretion, represented by actual responses of the monetary authority. The study compares their potential success in minimizing the gap between actual and desired state of the economy.

WAYNE K. TALLEY, Ph.D. Kentucky 1972. A model for distributing state aid to local school districts.

ROBERT D. TOLLEN, Ph.D. Texas (Austin) 1972. Bank asset management and the monetary policy lag.

The dissertation is an attempt to explain the behavior of the monetary policy lag through an investigation of the portfolio behavior of the banking system. The technique used is regression of the rate at which banks accommodate loans on the monetary lag in effect. This procedure requires generation of data reflecting the rate of change of loans, and extraction from this data-set of a behavioral variable for each observation over the business cycle. These behavioral variables represent the set of observations of loan accommodation which are regressed on the monetary lag.

YALCIN TUNCER, Ph.D. Pittsburgh 1972. Portfolio analysis with indirect utility.

This dissertation attempts to determine an optimal portfolio beyond the mean-variance analysis. The underlying rationale relates to the generalized risk aversion which depends upon functional forms of utility and existing moments of probability distributions. Using a specific utility function and the first four moments of an illustrative sample of common stocks, an empirical study is conducted. Given identical means and variances, the above aversion is found to imply a preference for positively skewed platykurtic portfolios.

TERRY L. TURNER, Ph.D. Carnegie-Mellon 1973. A theory of imperfect markets and of the demand for money.

THOMAS J. VALENTINE, Ph.D. Princeton 1972. An economic analysis of the Australian monetary sector.

This study describes a medium-sized quarterly econometric model of the Australian economy which was constructed with special emphasis on the monetary sector of the economy. Special attention is given to the behavior of the Australian equivalent of the U.S. commercial banks, the trading banks, and to the effect of this behavior on the real sector of the economy. The trading banks use the overdraft method of lending, and this factor is of central importance in the analysis. The final chapter of the study describes the complete model and presents equation estimates obtained by simultaneous equation techniques.

EDWARD E. VEAZEY, Ph.D. Michigan State 1973. Aggregation bias in the demand for money.

GEORGE M. VREDEVELD, Ph.D. Indiana 1973. Income redistributive effects of subsidizing public higher education in Indiana.

There has been a long tradition of subsidizing higher education in the United States. Yet, higher education subsidization policy suffers from a lack of serious economic review and analysis. The lack of analysis partly explains the situation in which educators agree that subsidizing education increases the quantity of education consumed but disagree about whether subsidizing education encourages a more equal distribution of income. This income-redistributive aspect of subsidizing higher education, which warrants serious study, is the focus of this dissertation.

DAVID R. WEINBERG, Ph.D. California (Berkeley) 1973. The effects of state sales taxation on prices, income, and employment.

BERNARD L. WEINSTEIN, Ph.D. Columbia 1973. The adequacy of the state tax base to support elementary and secondary education in 1980.

School enrollments and per pupil expenditures are projected by state to 1980 to determine the demands on the state fiscs. Market property values and personal income are projected by state to 1980 as measures of fiscal capacity. It is found that for thirty-seven of the fifty states, school costs expressed as a percentage of personal income should be lower in 1980 than in 1970. The case is made for full state financing of education.

JOHN A. WENNINGER, Ph.D. Washington (St. Louis) 1972. Political economy of monetary policy.

This research focuses on the limitations of conventionally defined monetary policy. Two types of problems complicate monetary policy with respect to the systematic attainment of the conventional objectives. These problems are structural constraints and the Federal Reserve's "support functions." A single model is incorporated to illustrate the possible implications of these "constraints" on the Federal Reserves behavior. This research results basically in a study of the Federal Reserves behavior from 1962 through 1971.

JOSEPH E. WHITBREAD, Ph.D. Syracuse 1973. A test of a total reserves based open market operations strategy for the Federal Reserve System, 1953–1969.

This is an empirical test of the feasibility of an open market strategy which employs total reserves as its near term target so as to control bank credit. The test was conducted over the years 1953–1969 using a monthly, non-linear model of the bank credit market patterned after the Brunner-Meltzer non-linear money supply hypothesis. It was found that gross national product was significantly influenced by the level of bank credit outstanding.

JOHN E. WHALLEY, Ph.D. Yale 1973. A numerical assessment of the April 1973 tax changes in the United Kingdom.

CHARLES E. ZECH, Ph.D. Notre Dame 1973. Regressivity and the property tax: Alleviation and tradeoffs.

This study is concerned with analyzing means for reducing the regressivity of the residential property tax. To this end, Minnesota state data are used to test two types of property tax relief programs: property exemptions, and the property tax credit. The property tax credit is shown to be more successful at alleviating regressivity at a reasonable cost to the state.

International Economics

JOHN M. ABBOTT, Ph.D. Tulane 1973. An investigation into the determinants of U.S. direct investment: 1950–68.

A simultaneous equation model of supply and demand for *U.S.* direct investment abroad is specified and estimated. The supply function includes the rate of return on domestic investment, the rate of return on foreign investment, and changes in the domestic flow of savings. The demand for direct investment is derived from a portfolio stock adjustment model. The model is applied to several countries. In general, changes in the level of *GNP* in the foreign country and changes in *U.S.* domestic savings are the dominant explanatory variables.

BIJAN B. AGHEVLI, Ph.D. Brown 1973. A model of the balance of payments and the money supply under the Gold Standard Regime: United States 1879–1914.

A theoretical open model is developed and tested econometrically to investigate the interaction between the real and monetary sectors. It is shown that contrary to the Cagan, Friedman, and Schwartz position, there were strong feedbacks from the real sector affecting the money supply through the balance of payments. The capital flows are shown to move procyclically (verifying Mundell's and Williamson's position), even though they were dominated by the anticyclical movements of the trade balance.

MALCOLM D. BALE, Ph.D. Wisconsin (Madison) 1973. Adjustment to freer trade: An analysis of the adjustment assistance provisions of the Trade Expansion Act of 1962.

HASSAN O. BALKHY, Ph.D. New Mexico 1973. International trade and income distribution: A 3x3 model.

A system is devised to rank commodities according to factor intensities when three factors are employed to produce three commodities. Some of the results rigorously analyzed in the 2x2 model are extended to the 3x3 model. The thesis employs the 3x3 model to analyze the cost of income redistribution when international trade tools are employed for that purpose.

BAROUCH BEN-ZION, Ph.D. Oregon 1973. Alternative concepts of export diversification: An application to the analysis of export growth.

TAPAN K. BISWAS, Ph.D. Rochester 1973. Nonshiftable capital and imported inputs in economic growth and international trade.

Economic growth in a trading economy has been discussed in the context of a nonshiftable capital structure. I have examined the significance of the myopic decision rule and have proved the stability of the long-run equilibrium when such a rule for investment allocation is adopted. Lastly, I have examined the role of imported inputs in a trading economy where capital is not shiftable.

ARYEH BLUMBERG, Ph.D. Chicago 1973. Eurodollars and the balance of payments: Some empirical evidence on two types of fractional Euro-dollar effect on the official settlements balance of the United States.

The thesis develops a quantitative framework for monetary approach to balance of payments. Demonstrates role of Euro-dollar reserve ratio by estimating fractional slopes of Official Settlements functions vis-à-vis Canada, Japan, eight European countries, using difference between Euro-dollar credit creation and Euro-dollar creation, termed conversions, as the independent variable. I demonstrate that in European case true Official Settlements balance is indeterminate within fractional range because of Euro-dollar deposits of monetary authorities.

ROBERT W. BOATLER, Ph.D. Cornell 1973. Trade theory predictions and the growth of Mexico's manufactured exports.

ROGER E. BOVE, Ph.D. Harvard 1973. Programming foreign aid with diminishing returns to investment.

This is a study of Chenery's two-gap model using a variable incremental capital-output ratio and a non-linear objective function. Evidence is presented to show that the usual assumption of a constant incremental capital output ratio is untenable. A capital accumulation function due to Robert Dorfman is introduced and explained. Finally non-linear programming the models for a 22-year plan are solved by Abadie's *GRG* method. Unlike its linear predecessors the model yields multiple-peaked capital flows but the three results due to Chenery and MacEwan stand.

I. BRYAN, Ph.D. Alberta 1972. Essays on transportation and trade.

THAI VAN CAN, Ph.D. Michigan State 1972. Capital flows and international adjustment.

This thesis uses a one-country model similar to Patinkin's as well as a two-country model to investigate the effectiveness of fiscal and monetary policy (consisting of changing outside and inside money) under the hypothesis that international capital flows influence aggregate demand by changing the level of real balances. Fiscal and monetary policy are generally found to be effective in each of the models.

FRANCISCO R. CASAS, Ph.D. Western Ontario 1973. Intermediate goods and the theory of international trade.

A survey of the recent trade literature reveals three major interpretations of the nature of intermediate goods: interindustry flows; pure intermediate products; and imported inputs. The thesis brings these interpretations closer to each other and provides some insights into the theory of intermediate goods in the context of international trade by investigating the general equilibrium properties of various two-sector models incorporating such goods.

DONALD M. CHAFFEE, JR., Ph.D. California (Davis) 1972. A contribution to customs union theory.

JOHN W. CHAMBERS, Ph.D. Claremont School 1972. Britain's balance of trade and the 1967 devaluation: A case study of the efficacy of devaluation in correcting a trade deficit.

The devaluation of the pound in November 1967 was accompanied by a number of other policy changes designed to reduce domestic demand (absorption), and hence contribute to an improvement in the balance of trade. This study is an attempt to isolate the impact of the devaluation, and to judge its effectiveness in improving Britain's balance of trade.

CHI-MING CHENG, Ph.D. Yale 1973. The determinants of international trade of manufactures: An empirical study.

ESHETU CHOLE, Ph.D. Syracuse 1973. Import substitution in practice: The case of the Ethiopian textile industry.

Import substitution has been the major strategy of industrialization in many underdeveloped countries. During the initial phase imports are curtailed (generally by a tariff wall) and production is geared toward satisfying domestic demand. When this is achieved there is need to move into the second phase—production of capital and intermediate goods, or production for export, or both. This paper attempts to assess various effects of import substitution in the Ethiopian textile industry.

YEI-CHOU BETTY CHU, Ph.D. Notre Dame 1973. Variable factor supplies and the pure theory of international trade.

The traditional fixed factor supplies assumption is abandoned. With the assumption of variable factor supplies and newly designed graphical analysis, various interesting points are found. Some findings are paradoxical in the light of ordinary economic theory. One of these is that output of a good responds perversely to its price. Another surprising contribution concerns specialization behavior in trade theory, namely, the country might go back to incomplete specialization after reaching complete specialization under certain conditions.

DAVID S. CLEAVER, Ph.D. Illinois (Urbana) 1972. A mathematical model of devaluation and relative price changes applied to Denmark in 1967–1968.

SCOTT F. CLOW, Ph.D. Illinois (Urbana) 1972. Trade and growth: A test of alternative theories.

FRANCIS J. COLELLA, Ph.D. Fordham 1973. An econometric study of United States-Canada short-term capital flows.

This dissertation measures the sensitivity of the short-term capital flow (between the United States and Canada, 1959–70) to the forces charged with its determination by the theory of short-term capital flows. These forces include interest rate differential, portfolios, money stocks, investment, prices, trade balance, and fiscal policy. Specification of the relationship between each variable and short-term capital flow is established by a review of the literature of theoretical and empirical studies.

RALPH K. COWAN, Ph.D. Michigan 1972. Effects of the United States-Canadian automotive agreement in Canada's manufacturing trade and price program.

WILLIAM P. CULBERTSON, JR., Ph.D. Texas (Austin) 1972. The burden of balance-of-payments adjustment.

This dissertation attempts to accomplish three goals. First, I survey the literature on the distribution of the international "burden" of balance-of-payments adjustment. Second, the relevant features of this literature are subjected to critical scrutiny, and the validity of the composite view of adjustment burdens is assessed. Finally, I find that the burden of international adjustment is always "shared" to some degree by the various nations involved. I conclude that the incidence has been largely visited upon the major surplus countries.

JAIME R. DAREMBLUM, Ph.D. Fletcher School 1972. Proposals for a general increase in the price of gold: An analysis.

GEORGE E. DRAKOS, Ph.D. Rochester 1973. Policy and the balance of payments.

The first part of this dissertation explores the effects on the balance of payments of the substitution of an equivalent value-added tax for the corporate income tax using a 2x2 and a 3x2 model of international trade for a small country. The second part uses an input-output model and investigates the price changes, following the tax change, and the resulting change in the balance of payments using Greek data.

ROBERT A. ELSON, Ph.D. Columbia 1973. Exchange rate policy and the performance of traditional exports in Argentina, 1946–1970.

This study shows the effects of changes in exchange rate policy on traditional exports in Argentina during the postwar period, before and after 1955. During the first phase, when traditional exports stagnated, exchange rate policy overvalued the Argentine peso vis-à-vis other currencies. However, during the second phase, when exchange rate adjustment was frequently used, traditional exports recovered. Theoretical and empirical results are presented in terms of a model of traded and nontraded goods.

NABIL C. M. FALTAS, Ph.D. California (Berkeley) 1972. Methodological and implementational issues in the design of international economic relations.

JERRY A. FEDELER, Ph.D. Iowa State 1972. An analysis of commodity aid and policies to eliminate its negative effects upon the commercial market.

Gross negative effects of foreign aid commodities are analyzed when aid displaces the recipient's production. An input-output model is used to show the distribution of negative effects among sectors of the Indian economy. Four policy alternatives to eliminate the losses in income arising from the negative effects are considered. One is a new application of a differentiated market. Monetary effects of concessional sales for the recipient's currency are shown to be an empirical question.

WARREN L. FISHER, Ph.D. Connecticut 1973. Processes in line, American agriculture, and the Leontief Paradox.

Following a thorough review of the literature of the Leontief Paradox, the Paradox is reexamined in light of Georgescu-Roegen's recent observation that parallel processes (such as agriculture) are overcapitalized relative to in-line processes. The upshot is that Leontief's estimates could have been biased by the existence of parallel-process industries in the United States. A method for estimating this bias is developed; data limitations prevented estimates.

ALLEN B. FRANKEL, Ph.D. Oregon 1973. A borrower-oriented model of U.S.-Canadian long-term fixed-interest capital flows.

JOHN H. GAILLIOT, Ph.D. Carnegie-Mellon 1973. Long-run determinants of the distribution of international monetary reserves.

DAVID H. GEDDES, D.B.A. George Washington 1972. U.S. direct investment overseas, 1968–1970: An analysis of corporate financial decision making as affected by the U.S. foreign direct investment program.

This dissertation establishes the optimum use of program allowables, the level of program constraint, and the minimum amount of long-term foreign borrowing use required to comply with the provisions of the Foreign Direct Investments Program for the years 1968–70 Computer modeling techniques are used, and individual company model results are statistically measured against actual company results.

KATHIE S. GILBERT, Ph.D. Louisiana State 1972. An evaluation of the association between Greece and the European community.

MANUEL GUITIAN, Ph.D. Chicago 1973. Devaluation, monetary policy, and the balance of payments.

A general equilibrium model shows how the government, by using aggregated demand policies, temporarily influences economic activity. In the long run, however, aggregate supply determines output and employment levels. International reserves are lost (gained) during the adjustment process due to policies which keep aggregate demand above (below) its long-run sustainable level. Such policies, if maintained, induce expectations of exchange rate changes. The self-fulfilling character of these expectations will force the government to devalue or revalue, thus neutralizing its previous demand policies.

JOHN J HAMPTON, D.B.A. George Washington 1971. Currency appraisal aspects of Euro-bond financing.

This study examines the factors which affect the costs of borrowing in different currencies for Euro-bond financing. It includes a background study of the Euro-bond market, a review of currency appraisal literature, a study of forecasting future exchange-rate adjustments, and a computer-based model for measuring the effective cost of Euro-bond financing.

RICHARD J. HERRING, Ph.D. Princeton 1973. International financial integration: Capital flows and interest rate relationships among six industrial nations.

Starting with the expected value maximization of a negative exponential utility function, equations for a two-country financial balance sheet are derived and the implications for capital flow and interest rate equations are deduced. Capital flow and interest rate equations are estimated for Canada, Germany, Italy, Japan, the United Kingdom, and the United States. Finally these empirical results are put together to construct an interdependent model of capital flows and interest rates.

MAXWELL W. HUDGINS, JR., Ph.D. Virginia 1973. Currency exchange rate fluctuations under a system of stabilization funds: The case of the pound-dollar rate, 1931–1938.

This study uses least squares regression analysis to investigate the fluctuations in the pound-dollar exchange rate from 1931 to 1938, emphasizing the intervention activities of the British Exchange Equalisation Account and the American Stabilization Fund. It is concluded that only the British Account had a significant effect on this rate. I also analyze whether or not competitive devaluations were pervasive in the 1930's and the implications of the Tripartite Currency and Gold Agreements.

LACY HUNT, Ph.D. Temple 1969. The Euro-dollar market capital flows, and the U.S. balance of payments: An econometric study.

ROBERT R. JOHNSON, Ph.D. Oregon 1973. Uncertainty in a general equilibrium model of international trade.

FRED R. KAEN, Ph.D. Michigan 1972. Euro-currency and national money market integration: An appraisal.

HARALD KNUDSEN, Ph.D. Oregon 1973. Expropriation of foreign private investments in Latin America.

ODIN K. KNUDSEN, Ph.D. Stanford 1972. A permanent income theory of export instability.

This study develops the concept of income instability as the dispersion of transitory income and utilizes the permanent income hypothesis of consumption to graft the study of export instability to the main body of economic theory. It derives a specific theory which incorporates a behavioral hypothesis from which to develop indices and empirical tests. The finding is that instability stimulates certain economic parameters in a direction predicted by the permanent income theory, a finding contrary to the a priori reasoning of extant literature on export instability.

RISHI KUMAR, Ph.D. Wayne State 1972. U.S. monetary and fiscal policies and the balance of payments since 1952: An evaluation.

GEORGE H. LAMSON, Ph.D. Northwestern 1970. Central banking in an open export economy: The monetary system of French-speaking West Africa, 1956–1967.

The goals of this dissertation were: to analyze a relatively exotic monetary union in which the central bank's major role was determining the allocation of bank credit through rationing at the rediscount window; and to better define optimum monetary policy for open, export economies. Its conclusions were that in such economies the money stock was better determined by a specie-flow-type mechanism than by discretionary counter-cyclical policies, and that "monetary independence" was limited to assuring that secular increases in the demand for money do not produce excessive official capital exports and that its counterpart savings are optimally allocated between competing domestic projects.

DANNY M. LEIPZIGER, Ph.D. Brown 1973. Long-term capital movements and economic policy: Canada (1953–61).

The dissertation examines the effects of foreign portfolio and direct investment on the Canadian economy under a regime of flexible exchange rates for the 1953–61 period. The thesis is further concerned with the respective roles of fiscal and monetary policy in the stabilization process, in cases of imperfectly mobile and perfectly mobile capital. A theoretical two-sector model is developed, based on five dynamic markets, proven stable and examined in equilibrium. The model is estimated by two-stage least squares.

LUIS LIBERMAN, Ph.D. Illinois (Urbana) 1972. Foreign trade and economic growth: The case of Costa Rica.

JOHANNES F. LIN, Ph.D. Cornell 1973. Economic policy in an open economy.

JENS R. LINDEMANN, Ph.D. Kentucky 1972. West German demand for foreign exchange reserves.

RACHEL MCCULLOCH, Ph.D. Chicago 1973. Import quotas and resource allocation.

The study compares quotas with tariffs as means of achieving national policy objectives. In some cases a quota is shown to entail a smaller welfare cost than a comparable tariff (where comparability is defined in terms of the policy objective underlying restriction). The relative costs depend on the policy objective, the structure of the market in the protected good, and the method of allocating the import licenses used to implement the quota.

MICHAEL S. MALONE, Ph.D. Kansas 1973. Tariff discrimination in a large country trade model: an application to customs union.

The study describes a three-country, two-commodity neoclassical model of international exchange in which one country can affect its terms of trade through the imposition of a tariff. The analysis of the model suggests that tariff optimization leads to discriminatory tariff rates in conjunction with customs cooperation.

JAMES L. MILLS, Ph.D. Oregon 1973. Analysis of the Mexican monetary system.

RAO S. MUSUNURU, Ph.D. Chicago 1973. Protection to fertilizer industry and its impact on Indian agriculture.

Effective rates of protection to the nitrogen fertilizer industry in India are estimated for the two periods 1961–65 and 1968–71. Production costs of such protection are computed using the effective rates. Consumption costs of protection are computed on the basis of estimated demand elasticities for fertilizer, and the net loss in agricultural production due to protection. These costs are shown to be significant, and

this protection may prove to be a serious drag on Indian agriculture.

GARY V. NOREIKO, Ph.D. Southern California 1972. A portfolio approach to domestic and foreign investment.

Along with the estimation of rates of return traditional mean-variance portfolio analysis is modified and extended to consider skewness of distributions of rates of return. In applications to foreign and domestic investment the sophisticated three dimensional portfolio model of rate of return, variance, and skewness is found to generate efficient portfolios that give a better explanation than those from traditional mean-variance portfolio analysis.

FERNANDO J. OSSA, Ph.D. Cornell 1973. Payments problems in the economic integration of Latin America.

CONRAD M. OULETTE, Ph.D. Clark 1972. The theory and practice of ERP Counterpart Funds in the United Kingdom, 1948–1951.

Examined are the roles played by *ERP* local currency, together with the rationales given to Congress for the use of these funds when it was considering legislation for the *ERP* and predecessor aid programs. It is concluded that the principal use of these funds in the United Kingdom—to retire privately held government short-term debt—was in the main economically insubstantive, but local currency was used substantively in Germany to establish financial intermediaries.

ANDREW H. PARNES, Ph.D. Stanford 1973. Export instability and the direction of trade.

This study looks at individual trade flows from one economy to another to find the degree and causes of export instability in a sample of developed and less developed countries. Structural variable categories are composition of exports, commodity concentration of exports, and direction of trade. Additional variables measure development level of trade partners and value of trade flow. Using covariance analysis, the model indicates that less developed nations experience significantly higher export instability than developed countries and that choice of trading partners influences level of instability. Significant variables are level of commodity concentration and magnitude of trade flow. Higher concentration levels produce greater instability.

LORENZO L. PEREZ, Ph.D. Pennsylvania 1972. The export performance of the Latin American countries during the 1960's.

HOWARD C. PETITH, Ph.D. Rochester 1973. Joint production and the theory of international trade.

By careful consideration of the character of natural resources, it is shown that received theory implies that not more than 5 goods may enter trade, while actually over 30,000 do. A new model of trade is constructed based upon joint production and nontraded goods which alleviates this absurdity, and leads to a number of unorthodox and fairly robust results regarding the relative prices of nontraded goods and factors and the pattern of trade.

JONATHAN J. PINCUS, Ph.D. Stanford 1972. A positive theory of tariff formation applied to nineteenth century United States.

The effects of an import duty are nonexclusive, so I adapt public goods supply theory to explain the industrial pattern of tariffs. The focus is on appropriability of quasi rents, and on bargaining and transaction costs. I reinterpret the infant industry tariff phenomenon. An *OLS* study suggests it was a disadvantage in 1824, to be represented by few congressmen and senators.

ROBERT O. REID, Ph.D. Georgetown 1972. Border tax adjustments and international trade flexibility.

The thesis analyzes the effect of border tax adjustments on trade flows under various monetary and fiscal policies. A model was developed and used to estimate the impact of border tax changes on the trade balances of Western industrialized countries. The use of border taxes as a policy instrument was also considered.

HAK Y. RHEE, Ph.D. Kansas 1972. Optimal external debt for a developing economy.

The dissertation deals with a small country's optimum foreign debt over a finite planning horizon. The problem is analyzed in an economy with multi-outputs and multi-inputs. The input vector is divided into two subvectors, one of them being the domestic inputs and the other being the foreign inputs. The outcome of the analysis shows the condition under which the additional debt should be zero as well as other efficiency conditions along the optimum path.

MIGUEL RODRIGUEZ-VILLENAVE, Ph.D. Oregon 1972. Import substitution: The case of Venezuela.

GEORGE SAMUELS, Ph.D. Oregon 1972. The role of imports in growth: Selected Latin American countries, 1953–1967.

EDMUNDO SANCHEZ-AGUILAR, D.B.A. Harvard 1973. The international activities of U.S. commercial banks: A case study, Mexico.

The role of *U.S.* commercial banks in Mexico is examined. The volume of *U.S.* banks' claims on Mexico is estimated to be around $5,000 million. A close look at *U.S.* commercial banks' methods of operation abroad is presented, including the sources and profitability of foreign lending. Chraracteristics of *U.S.* bank loans to Mexico are enumerated. Finally, Mexico's ability to repay her external debts, present and future, is discussed in detail. Strong recommendations are made to the parties involved.

LUCIO P. SCANDIZZO, Ph.D. California (Berkeley) 1972. International capital movement and economic growth.

Economic growth and international capital movements are examined using a neoclassical model and in the second stage a dynamic post-Keynesian "two-gap" model. Trade and capital movements between a less developed country and a capital rich country with monopoly power are analyzed. Under conditions of a dominant trade gap, optimum paths of foreign exchange inflow are determined with the results applied to case studies of Argentina and India.

RICHARD W. SCHWINDT, Ph.D. California (Berkeley) 1973. The effects of economic integration upon industrial structures: The case of France, Italy, and West Germany.

GRANT M. SCOBIE, Ph.D. North Carolina State 1973. The effect of changes in the protective structure on the exchange rate: An analysis of the New Zealand foreign exchange market.

The underlying premise is that the exchange rate and the level of protection are a related pair of policy instruments for achieving external balance. The devaluation rate and the welfare gains associated with variable reductions in protection are estimated. The model is used to estimate the impact of British entry to the EEC on the New Zealand foreign exchange market.

KARNAIL SINGH, Ph.D. Western Ontario 1973. A monetary view of the balance of payments: An empirical investigation.

The empirical investigations reveal that in all the fourteen industrialized countries the balance-of-payments surplus is a positive function of the "sterilized" excess demand for money in the world as a whole; the extent of *U.S.* short-term liabilities to foreign central banks is a "system determined" phenomena; and the country demand for money functions are stable.

MILTON D. STEWART, JR., Ph.D. Texas (Austin) 1972. Foreign direct investments in the United States.

This study examines selected aspects of foreign direct investments in the United States, an area that has been apparently relatively ignored by scholars, general researchers, and popularizers alike.

ANDREW STOLLAR, Ph.D. Boston College 1973. Cotton textiles: An empirical evaluation of voluntary export restraints.

An attempt is made to evaluate the effects of the GATT sponsored long term arrangement regarding international trade in cotton textiles during two of the initial years of the arrangement (1964 and 1965). A spatial equilibrium model of international trade is used to simulate trade flows (and domestic consumption and production levels). The empirical results indicate that *LTA* (Long Term Arrangement) has had a restrictive influence on the exports of selected major Asian exporting countries.

SUK-TAI SUH, Ph.D. Washington (St. Louis) 1972. A general model of factor proportions theory of trade.

The dissertation is concerned with the following: to generalize the simple factor proportions theory in the case where commodities outnumber factors; to construct a new production function which is neoclassical in form but contains human capital as a third factor of production; to establish a new comparative advantage criterion in terms of the composite capital-labor share intensity; and to test the above for Korea, Japan, and the United States.

DAVID M. SWANSON, Ph.D. North Carolina (Chapel Hill) 1972. The demand for money and the balance of payments.

SUBHASH M. THAKUR, Ph.D. California (San Diego) 1972. Theory and estimation of speculative capital movements.

The traditional theory of forward exchange market and international capital movements is reviewed. A model of speculation when the currency is under speculative pressure is developed. A stock-adjustment model is then estimated for the United States using data for the period before the 1971 devaluation of the dollar. Impact of speculation on *U.S.* claims and liabilities is found to be substantial during this period.

MAHMOUD S. H. TOUBAR, Ph.D. Pittsburgh 1971. Elasticities of export supply in international trade.

JOSEF VAN BRABANT, Ph.D. Yale 1973. Bilateralism and intra-CMEA trade: 1949–1966.

EWEN M. WILSON, Ph.D. North Carolina State 1973. An elasticity analysis of the Rhodesian trade embargo.

An elasticity framework is developed to predict the welfare and exchange rate effects of the Rhodesian trade embargo. The notion of the uniform tariff equivalent is adopted to estimate price effects of the trade ban in the import and export markets. Explicit recognition is made of the importance of flue-cured tobacco in the Rhodesian economy.

Business Administration; including Business Finance and Investment, Insurance, Marketing, and Accounting

SURENDRA P. AGRAWAL, Ph.D. Florida 1973. The role of accounting in computer-based information systems.

This study presents a model of information systems that may overcome the deficiencies observed in the functioning of accounting and MIS. The proposed model would control its own input and output, provide for computerized as well as manual processing of data, and receive and utilize feedback from users, for its development. In addition to systems and O.R. specialists, it would include an expert in "information control" which may develop as a branch of accounting.

ERIC L. ANDERSON, Ph.D. Georgia State 1973. The buffer motivated production-to-stock decision and its macroeconomic implications.

This paper constructs a model to represent manufacturers' production-to-stock behavior with a sales buffer motive to determine the production flexibility in responding to unexpected sales. The model is tested on 6 two-digit S.I.C. industries that produce to stock—not to order. The resultant maximum likelihood estimators are then inserted in the basic Keynesian aggregate demand function to determine the macrodynamic implications of their behavior.

HENRY D. ANGELINO, D.B.A. George Washington 1973. An analysis of the U.S. Army's system for developing estimates of major equipment investment costs.

This research is a critical examination of the U.S. Army's cost estimating function for developing estimates of major equipment investment costs. The research supports conclusions that cost estimates recently prepared by the Army are timely, accurate, and complete. However, estimates have tended to be understated. Members of the cost community hold that staffing, organization, and operation of the cost function, facilitate preparation of accurate estimates.

ROGER M. ATHERTON, Ph.D. Michigan 1972. The impacts of centralization on performance and supervisory perceptions of centralization, attitudes, behavior, and effectiveness.

EMIL ATTANASI, Ph.D. Missouri 1972. Uncertainty in road construction: Implications for state highway agencies.

This study relates firm bid pricing behavior to industry bid distributions. Analysis of sequential bid pricing is conducted under behavioral assumptions identified with expected mark-up maximization and utility maximization. The influence of alternative regulatory policies, bidder risk attitudes and resource uncertainty is analytically examined. Properties of bid distributions are analytically examined and related to economic structure of the market, attitudes toward risk, and expectations of bidders. Bid data are then examined and provide a basis for comments associated with applied research in this area.

BRUCE N. BAKER, D.P.A. George Washington 1972. Improving cost estimating and analysis in DOD and NASA.

Surveys were conducted to determine cost estimating and analysis approaches and preferences within the Army Materiel Command, Air Force Systems Command, Naval Materiel Command, and National Aeronautics and Space Administration. It was found that most respondents distrust the parametric approach and have more confidence in detailed engineering approaches to cost estimating. Respondents called for several improvements in cost estimating but disagreed, on average, that 1970 procurement will result in less cost growth.

JIMMY D. BARNES, Ph.D. Oregon 1973. A strategic competitive bidding approach to pricing problems for petroleum industry drilling contracts.

ROBERT R. BELL, Ph.D. Florida 1972. The potential of pupillary response in business research: An investigation of methodology and autonomic contamination.

The debate over the use of pupil changes as indicators of physchological response cannot be resolved until methodological shortcomings are overcome. This study evaluated the validity of pupillometric technology, using real-time computer sampling of six physiological variables in a totally automated measurement system. Cost effectiveness studies of computerized methodology were conducted. Analysis showed that while pupillometric may be limited in general usefulness, computerized controls and measurements present advances in methodology, and can be economically justified.

URI BEN-ZION, Ph.D. Chicago 1972. Measures of risk in the stock market and the valuation of corporate stock.

The first part of the thesis analyzes the rate of return framework and the valuation framework of testing the "riskiness" of alternative risk measures. The second part analyzes the effects of investment in tangible assets, such as research and development and advertising, on the valuation of corporate stocks. The empirical work is done in a cross-section study, but it also analyzes a time-series measure of risk.

WILLIAM J. BRENNAN, Ph.D. Michigan 1972. Investment analysis and generally accepted accounting principles.

DOUGLAS B. BROWN, Ph.D. Michigan 1973. Modelling of retail patronage promotions—the practice and its value to management.

RICHARD W. CALLEN, Ph.D. Texas A&M 1972. A dynamic investment, production and financial model for a competitive firm.

In this dissertation, a discrete time control for production, investment, and finance is formulated for a perfectly competitive firm in a certainty world. The usual production function is not used. Instead, a linear process type of function is adopted. The model assumes linearity among the capacity variables. With these assumptions, an operational model for a multiproduct firm with multi-capacities under a form of capital rationing is developed. The firm is assumed to maximize its net wealth.

JOHN C. CAMILLUS, D.B.A. Harvard 1972. Formal planning systems: The control considerations in design.

The study examined the design implications of the purpose of facilitating the control of operations through formal planning. The basic proposition was that, in fulfilling this purpose, another important pur-

pose—namely, ensuring the stimulation and exercise of executive creativity—is adversely affected. An approach to the design of a segment of the corporate planning system, based on the proposed control/creativity tradeoff was developed and then revised on the basis of a survey of practices in over eighty corporations and interviews with practicing executives.

GEOFFREY O. CARLINER, Ph.D. California (Berkeley) 1972. Headship and ownership.

DARYL E. CARLSON, Ph.D. California (Berkeley) 1972. Financial analysis of institutions of higher education for policy decisions.

TING KO CHEN, Ph.D. Michigan 1973. Management transfer, management practice, and management performance—an empirical quantitative study in Taiwan.

PETER S. CHUNG, Ph.D. Northwestern 1972. Measurement and explanation of expected earnings to price ratios of common stock.

The major interest and focus of this dissertation is how to measure correctly "expected earnings" which appears in the numerator of the rate of return on common stocks. Four methods of measuring expected earnings were employed in the cross-sectional multiple regression analysis to explain different rates of return among the sample firms. The results show that the four methods provide better explanation for differential rates of return on common stocks than any other existing models.

JAMES J. CLARKIN, D.B.A. Harvard 1973. An investigation of a performance evaluation process.

This study examines data from performance evaluation records and from two personality tests with a questionnaire administered to 97 employees of a large government organization. The scores correlated closely with promotion decisions indicating reliance on minute differences. Better performers were distinguished from poorer only by the need to impress others, where they were brought up, and their number of siblings. There was dissatisfaction with communications about performance, and poorer performers felt more negative about their experiences and the promotion system in use. Poorer performers were also less self-confident and more negative about their future.

PHILLIP M. COLEBATCH, D.B.A. Harvard 1973. Manufacturing policy in the multinational enterprise: A mathematical study of sourcing patterns and their relationship to corporate financial goals.

This thesis examines production allocation, sourcing patterns and transfer pricing in the multinational enterprise, and studies their interaction with each other and with financial tools such as management fees, royalties, and internal dividend policy. Specifically, a mathematical model of a multinational corporation is developed and used as a basis to study these interactions, and to study the differences in the optimal choice of these quantities for two different objectives, consolidated profits and cash flow to the parent.

RICHARD C. CUBA, D.B.A. George Washington 1971. Relationships between technologists' motivation to perform, role perception and participative management: A study of engineering/scientific manpower utilization in R&D-oriented companies in the Baltimore-Washington area.

This research investigates the effects of role perception and participative management on the job performance of 600 engineers and scientists in five firms. Specific workday utilization patterns were secured to provide a measurement of the large existing mismatch between skills, interests, and job requirements. Sizable pockets of role uncertainty and laissez-faire management in the technical areas were uncovered. Organizational role clarification and participative interaction were found to be significantly related to higher performance levels and motivation to perform. Role orientation was also found to be flexible with age, experience, and traditional organizational rewards.

DAVID J. CURRY, Ph.D. California (Berkeley) 1973. An empirical test of relations between an individual's preference and similarity space of brands.

JEAN M. DAVID, Ph.D. Florida State 1972. A model for the efficient allocation of resources in a state university.

This dissertation applies an activity analysis model to the allocation of resources in a state university. The problem is first formulated as a linear programming model which serves as a first approximation and makes possible the mapping of the university's production possibility frontier. The model is then formulated as a non-linear program. A stochastic element is later introduced into the non-linear program. This is followed by a dynamic formulation with varying horizons.

DAVID F. DIANICH, D.B.A. George Washington 1971. An application of non-linear programming to approximate optimal management control decisions.

The research investigates a technique for optimizing management science models characterized as dynamic, non-linear, and constrained. The research integrates a classical solution procedure, the method of Ritz, in variational calculus, with contemporary non-linear programming techniques. The unknown dynamic decision function is approximated by a function of n variables, an n-dimensional polynomial, tranforming the model to an approximation in n-dimensions. Contemporary techniques of non-linear programming are used to optimize the transformed, n-dimensional model.

BERNARD C. DILL, D.B.A. George Washington. A regional study of existing developments in and between financial policy and financial operations at the headquarters level of selected large corporations.

This study concerns developments in the business finance function at the headquarters level. Matters

discussed concern ways in which the importance of the chief financial executive and other elements of financial management along with the relationships between financial policy and financial operations are changing. Discussed are ways in which changes in this function's environment and in the development of information technology are influencing this function. Interviews were conducted among 32 firms from *Fortune's 500 Directory.*

WILLIAM A. DIMMA, D.B.A. Harvard 1973. The Canada Development Corporation: A diffident experiment on a large scale.

The Canada Development Corporation was proposed to the Canadian Parliament in 1963 as an instrument of both economic nationalism and of gap-filling industrial development. It would operate under both private and public-sector influences. By 1971 and enactment, it had swung more solidly into the private sector and behind industrial development, partly reflecting a pervasive business ethic. However, the enabling legislation does not inhibit a future government, under some circumstances, for using the *CDC* for other purposes.

WILLIAM A. EARNER, D.B.A. Harvard 1973. An assessment of organizational elements that affect the introduction of a computer-based simulation into some organizational decision processes.

Introducing a complex, computer-based model to an organization not only affects the rational decision processes of that organization, but also the organizational system in which those processes are embedded. This thesis examines in detail two organizational decision processes in a bank about to implement a computer-based environmental simulation. These processes are described in terms of the segments of the decision processes interacting with four relevant elements of the organizational system. The thesis concludes with an assessment and prediction of model implementation in light of the interaction of decision processes and organizational system.

JON ENGLISH, Ph.D. Florida 1972. The use of a multivariate organizational behavior model in describing individual, group, and intergroup behavior: A comparison of three manufacturing firms.

The primary objective of this study is to describe and analyze behavioral patterns in organizations in terms of structural and leadership variables. A secondary purpose is to employ a new multivariate "systems" model that explains human behavior within an organization and to assess the operational feasibility of the model. Data on structural, leadership, and behavioral factors in three manufacturing plants were gathered from personal interviews, observation, and questionnaires. Each plant was described and contrasts in behavioral patterns were analyzed.

WALLACE FELDMAN, D.B.A. Harvard 1973. A study of selective marketing communication.

This thesis investigates the potential of advertising aimed at individuals or small recipient groups. It examines the development and use of increased audience selectivity capabilities in print, mail, and air media. An assessment of present constraints and possible future developments in message production and delivery technologies is made. The applicability of market segmentation concepts to selective communication strategies is discussed. Finally, an experiment assesses the possibilities of increased effectiveness of segment-specific advertising compared to advertising generalized for a total audience.

MARCO R. FIORELLO, Ph.D. California (Berkeley) 1972. Management and design tools for document retrieval systems: A method for predicting quantity output.

RICHARD G. FISCHER, D.B.A. Harvard 1973. Improving investment performance and the role of fixed-income securities.

Efficient portfolios of bonds and stocks are constructed using the Markowitz/Sharpe technique. For wide ranges of market conditions and portfolio risk levels, bonds play a much greater portfolio role than is commonly perceived. A large fraction of the bonds serves as a surrogate for a riskless asset and is relatively insensitive to the market outlook while the remaining bond investments may comprise a substantial proportion of the risky asset portfolio depending only on the pertinent market forecasts.

ROBERT J. GAYTON, Ph.D. California (Berkeley) 1973. A critical examination of presently accepted international accounting translation techniques.

LOUIS GILL, Ph.D. Stanford 1973. Iterative multilevel planning with aggregation.

Procedures are developed for plan preparation though explicit introduction of aggregation-disaggregation methods in standard "decentralized" planning procedures. A central planning bureau deals with aggregated quantities (prices and quantities of groups of products). Regional or sectoral bureaus are responsible for aggregating and disaggregating information. Production units are responsible for their choice of technique; they deal with disaggregated quantities which are the fine details of the plan. The utility function is of the Kantorovitch type.

STUART C. GILMOUR, D.B.A. Harvard 1973. The divestment decision process.

In this study divestment decisions were found to be the personally painful and organizationally difficult outcomes of a high level political decision process that made limited use of economically rational arguments. The study used four well-known management decision process models to look at the events surrounding divestment decisions in three large business organizations. Using these analyses, a model of the divestment decision process was constructed and implications for management and research were drawn.

JOHN H. GRANT, D.B.A. Harvard 1972. Management implications of systems-oriented strategies within selected industrial firms: A developmental model.

This research had as its objective the formulation and evaluation of a developmental model for analyzing four primary variables associated with the management of systems-oriented operations. Field research executed in five large corporations revealed that strategic, market, and technological factors are interdependent with, not determinant of the organizational structures surrounding systems work. Forces influencing changes in systems strategies and related research topics are discussed.

STEVEN D. GROSSMAN, Ph.D. Tufts 1972. Alternative theories of inventory investment.

THOMAS D. HINTHORNE, Ph D. Oregon 1973. Perceptual differences as a constraint on the implementation of management science techniques.

GERARD G. JOHNSON, D.B.A. Harvard 1972. The role of formal evaluation in the process of budgetary control.

In addition to accounting for the financial results of operations, the evaluation process must begin to take into account the feelings of the individual manager. This research project demonstrated that an important aspect of the evaluation process is the preparation it provides for future performance. Under varying economic conditions laboratory subjects who were so prepared outperformed those evaluated under the more traditional accounting variance format.

L. TODD JOHNSON, Ph.D. Michigan 1972. The auditor's responsibility for the detection of management fraud: Selected cases.

JOSEPH E. KASPUTYS, D.B.A. Harvard 1972. The influence of government profit policies on contractor management.

The Department of Defense determines profit objectives for negotiated contracts largely as a percentage of costs, without recognition of contractor assets. This policy should encourage contractors to act contrary to government objectives by keeping assets low and costs high. This was tested through a comparative study of defense and commercial operations, two questionnaires and research into the British asset-based system. The findings confirmed that existing profit policies are dysfunctional and could be improved through some consideration of assets.

JOHN B. KAYE, D.B.A. George Washington 1973. A test of the square root rule of stock price movement.

The study tests the hypothesis that the relationship between the price movement of low-priced and high-priced stocks during market rises and declines is mathematically consistent. Although prices move in different amounts, the square roots of prices move in equal amounts, on the average. The study fills deficiencies and gaps of earlier studies and investigates a new dimension: expected growth rate. It has implications for the *alpha* and *beta* of modern portfolio selection.

DIETER KLEIN, Ph.D. California (Barkeley) 1973. Dynamic behavior of adjustment processes.

JAY E. KLOMPMAKER, Ph.D. Michigan 1973. Sales forecasting for new industrial products.

CHARLES LILLIS, Ph.D. Oregon 1973. An investigation of the uses of non-metric multi-dimensional scaling in the identification and analysis of market segments.

JAY D. LINDQUIST, Ph.D. Michigan 1973. Retail store image and its role in the purchase decision process: An empirical study.

BRADLEY D. LOCKEMAN, Ph.D. Michigan 1973. The consumer information seeking process for a fashion good—An exploratory study.

NORMA M. LOESER, D.B.A. George Washington 1971. An examination of the top-level leadership strategy emanating from the executive branch of the government having an impact on the military-industry interface.

The study is oriented toward obtaining an explicit understanding of the military-industry interface through examining: the leadership authority and power of the President as chief of foreign policy and as commander-in-chief; the roots of *U.S.* military commitments abroad; the national objectives having defense implications as ascribed to the President; and the extent to which federal budget outlays for national defense generated growth in the defense establishment and defense firms.

RONALD W. LUNDQUIST, Ph.D. Oregon 1973. Organization and the seasons: An exploratory study of the structural and behavioral effects of seasonal employment fluctuations in the food processing industry.

NEIL E. MCNEILL, D.B.A. Harvard 1973. The use of profit as a measure in a nonprofit organization.

The study addresses the use of a constructed profit measure to improve performance in a nonprofit organization. The study applied a profit measure to three nonprofit Navy commissary stores while using three similar stores as a control group. The conclusion is that the technique is potentially useful, however, its effectiveness is sensitive to the degree of congruency existing between the organizational structure and the organizational purpose.

WILLIAM G. MISTER, Ph.D. California (Berkeley) 1973. Temporal additivity: A problem of accounting measurement.

JEFFREY H. MOORE, Ph.D. California (Berkeley) 1973. Peak-load allocation of a time-shared computer in a quadratic team.

JAMES R. MORRIS, Ph.D. California (Berkeley) 1972. The decision to purchase information for portfolio choice.

GERALD L. MUSGRAVE, Ph.D. Michigan State 1973. Economies of scale in computer centers.

IKUJIRO NONAKA, Ph.D. California (Berkeley) 1972. Organization and market: Exploratory study of centralization and decentralization.

DAVID P. NORTON, D.B.A. Harvard 1973. The centralization of information systems in large organizations.

The objective of the research was to determine if there is a preferred manner of assigning responsibility for the computer and associated tasks in various organization environments. The research, which was based on intensive case studies in two multi-divisional organizations, showed that economically rational arguments exist for various degrees of centralization or decentralization in specific environments. The organization forms dictated by economically rational arguments are frequently ignored in practice. The research explores the reason for this inconsistency.

JAMES A. OHLSON, Ph.D. California (Berkeley) 1972. The usefulness of accounting concepts based on accruals: A Bayesian approach.

JOSEPH R. PISANI, Ph.D. Maryland 1972. Automotive repair: A study of consumer and repair shop attitudes toward the industry, the performance of repairwork and regulation.

The dissertation reports the findings of a mail survey of the motor-vehicle-owning and repair shop populations of Prince Georges County, Maryland and Travis County, Texas. A general background on the automotive repair problem and a review of current literature on the industry are provided. The overall quality of warranty and nonwarranty repairwork was found to be substandard. Consumers and repair shops strongly favored regulation of mechanics and/or repair shops at the state level.

ROLLA W. RIDER, Ph.D. Oregon 1973. The marketing executive: A profile.

SIMCHA SADAN, Ph.D. California (Berkeley) 1972. The firm-investor interface: An inquiry into the role of accounting information.

DOUGLAS F. SCHOFIELD, D.B.A. Harvard 1972. The planning process in institutions of higher education.

Starting with a traditional four stage view of a planning process (goals, objectives, budgeting, evaluation), two sets of variables are identified as useful in analyzing a college planning process and in implementing improvements. One set pertains to planning functions in each stage, and the other pertains to the dynamic, integrated functioning of the whole process. Planning in two colleges is examined, and recommendations for management actions are made. Planning implications of three major trends in colleges' environments are discussed.

HENRY H. SEWARD, D.B.A. Harvard 1973. Measuring user satisfaction to evaluate information systems.

This dissertation proposes that the evaluation and design of information systems should be considered separately. Evaluation, if separated, may be conducted by less expensive means than design. The approach suggested is to evaluate by measuring user satisfaction with the information system. Satisfaction is considered to be directly related to the number of times the information system does not contain required information and the user must conduct an information search. The approach was tried and found feasible.

MARK G. SIMKIN, Ph.D. California (Berkeley) 1972. Mathematical programming for the state earmarking process: Tax policy in theory and an application for highway finance.

CHARLES C. SNOW, Ph.D. California (Berkeley) 1972. Environmental uncertainty and organization behavior.

SUSAN SOLOMON, Ph.D. California (Berkeley) 1972. Determination of satisfactory scale for data processing facilities.

ROBERT E. STERNENBERG, Ph.D. Houston 1972. A study of full cost versus marginal cost pricing of proposals to government agencies and their impact on contract award and performance.

WILL E STRAVER, D.B.A. George Washington 1971. A study of attitudes toward organizational loyalty.

Survey compares attitudes between employees from business government, military. Findings also analyzed by age, position, rank, seniority. Chi-square analysis on thirty-four attitudinal statements. Few significant differences found at .95 confidence level. Main content of statements shows similarity between loyalty enhancers and motivational factors. Organizational loyalty ranked fourth behind family, self, country. Supervisor found to be major loyalty object in each type organization. Respondents strongly questioned value of loyalty oaths.

WALTER J. STUART, JR., D.B.A. George Washington 1971. An empirical assessment of the managerial utility of certain general systems concepts: The savings and loan association.

This study seeks to make certain general systems concepts as applied to management more defensible by defining a savings and loan association as a system and examining the behavior of a large number of as-

sociations during a period of unusual reduction in net inflow of savings. Using the empirical results obtained the study applies these concepts to the problems of financial intermediation, designing a General System of Financial Investment Flows.

FREDERICK W. TEBBEN, D.B.A. George Washington 1971. Cash tender offers in 1967: A study of management and stockholder experience.

Cash tender offers, or take-over bids, that started in 1967 were studied from the standpoint of premiums offered, relationships of market prices and tender prices to book values, movements in stock and earnings growth rates and dividend incomes. One of the most important conclusions was that the mergers that followed tender offers had unfavorable results for both companies and stockholders. Earnings growth rates and stock prices suffered.

JOHN F. TOMER, Ph.D. Rutgers 1973. Management consulting for private enterprise: A theoretical and empirical analysis of the contribution of management consultants to economic growth in the United States.

The basic contention is that management consultants, due to their work for clients, contribute in many ways to economic change, which makes possible and "explains" a certain amount of annual economic growth. The analysis indicates that the lasting contributions of these consultants are types of human capital or, perhaps, organizational capital. Estimates of this capital formation and the resulting aggregate annual economic growth are made. Interviews in thirty-two management consulting firms provide the empirical basis.

JERRY A. VISCIONE, Ph.D. Boston 1973. The introduction of growth into the traditional theory of capital structure.

JAMES A. WALKER, D.B.A. Harvard 1973. Using analysis effectively in a complex decision process-planning NASA's planetary programs.

The problems of using systems analysis techniques to influence the outcomes of policy making in a complex organizational and political context are illustrated by case studies of attempts to apply such methods to long range planning for NASA's unmanned planetary programs. A method of enhancing the effectiveness of analysis in such settings by means of self-conscious "process-oriented approach" utilizing organizational, political, and psychological models as conceptual tools is proposed and discussed.

KENNETH D. WALTERS, Ph.D. California (Berkeley) 1972. Freedom of speech in the modern corporation.

BERNARD S. WATERMAN, D.B.A. George Washington 1971. Small business participation in federal government research and development—the government perception.

This study examines the attitudes of government technical and procurement personnel purchasing research and development (R&D) toward participation of small business in government R&D. It develops the views of these personnel regarding the availability and interest of small firms, potential measures of assistance, and small business performance. The government R&D purchasing community believes that small business lacks sufficient technical capability and does not mount a sufficiently aggressive marketing effort.

DAVID P. WEINER, Ph.D. Michigan 1972. The feasibility of obtaining an objective measure of the current value of land and buildings for disclosure in published financial statements.

ARMAND B. WEISS, D.B.A. George Washington 1971. Development of guidelines for the determination of economic retention stock in the federal government.

Six economic retention stock models are developed and compared with two other models and to *ERS* policies of the Defense Supply Agency and the General Services Administration. The *DSA* and *GSA* policies are found to be not economic. Six types of sensitivity analysis are made. The method of determining total quantity of stock to hold is derived. Decision tables are provided to cover a variety of situations.

RICHARD A. WERBEL, Ph.D. Michigan 1972. A multivariate analysis of brand loyalty for purchasers of new cars and major household appliances.

LAWRENCE M. K. WONG, Ph.D. Michigan 1973. Managerial ideologies in industrial society: The construction of the corporate world-view.

EARL M. WYSONG, JR., D.B.A. George Washington 1972. An analysis of the effectiveness of computer-based financial management systems in the civil agencies of the federal government.

The dissertation demonstrates: the complexity of requirements for federal financial management systems; some computer concepts available to handle the information requirements; and how federal agencies are handling the situation. Through the use of questionnaires, it was found that the capabilities of the computer have not been employed to a satisfactory extent. Although federal agencies' accounting systems provide the required financial information for external reporting purposes, they are deficient in providing information for managerial purposes.

CHUL KOO YUN, Ph.D. Michigan 1973. Transfer of American management know-how to foreign environments: Applying managerial budgeting system in Korea.

SIMON J. ZYSMAN, D.B.A. Harvard 1973. Top management and decentralized investment planning in diversified firms: A comparative study.

This dissertation is a comparative study of the top management role in major investment plans of two business units in each of three highly diversified *U.S.*

corporations of $500 million to $1 billion sales. The important variables selected to study are: the key participants in the investment planning process; interactions among the participants; the information available to them; and the organizational framework of the process.

Industrial Organization and Public Policy; including Economics of Technological Change, and Industry Studies

NISAR AHMAD, Ph.D. Washington State 1973. Economic utilization of road and rail transportation in Pakistan.

This thesis involves an attempt to examine Pakistan's transport pricing, regulatory, and investment policies from the point of view of economic efficiency. It found considerable evidence that the scarce resources devoted to transportation in West Pakistan were not utilized efficiently. Misallocations are found in passenger traffic and freight traffic services, and the pace of introduction of new transport technologies. A key area of misallocation pertained to the diversion of high-value freight traffic from rail to road transport.

STEVEN ANDERSEN, Ph.D. Rutgers 1972. Regulation, supply and cost in air transport: The domestic trunkline industry.

Scheduling behavior is examined under the assumption of imperfect cartelization through regulation. The model yields a functional relationship between load factor and fare, production cost, and a set of exogenous, demand determining variables. Estimation results, supplemented by an analysis of adjustment lags for costs of production, are used to evaluate changes in regulatory policy adopted during 1970, with particular emphasis on the tradeoff between fares, load factor, and rate of return.

PAUL R. AUERBACH, Ph.D. Wisconsin (Madison) 1973. Financial institutions and corporate control.

ROBERT BARRASS, Ph.D. Washington State 1973. The pre-1963 marketing of automobile tires by gasoline retailers: A problem of market power generated by franchise agreements.

This thesis investigates the implications of restrictions imposed by oil companies on the selection of brands of tires, batteries, and accessories by their franchised dealers. Franchising links the enterprise of small proprietors with the widely accepted brand names of large corporations, combining advantages of big and small business. Many restrictions are placed on franchisees and their purpose is brand protection, but these restrictions implement hidden surcharges, facilitate price discrimination, or extend the franchisor's monopoly power to additional lines.

JOHN H. BEEBE, Ph.D. Stanford 1972. Institutional structure and program choices in television and cable television markets.

The thesis constructs a model of product competition in television. It extends models by Steiner, Rothenberg, Wiles, and McGowan, and uses theories of spatial competition and public goods. The model compares program patterns of monopoly and competition under advertiser and subscriber support. Predictions refute Steiner's result that under limited channels and advertiser support monopoly provides greater diversity than competition, and shows the opposite for unlimited channels. Furthermore, pay TV promises programs closer to viewer preferences.

NAROTTAM BHARGAVA, Ph.D. Pennsylvania 1973. The impact of organization form on the firm: Experience of 1920–70.

RONALD S. BOND, Ph.D. Indiana 1972. Aggregate economic structure: Mobility, concentration, and mergers.

This dissertation is an economic and statistical analysis of aggregate mobility. The study of aggregate mobility reveals differences in the relative growth of the largest firms and provides a vehicle for analyzing the dynamics of aggregate concentration. Because past studies of aggregate mobility have been interpreted more often in a market than in an aggregate context, the relationship between aggregate mobility and changing aggregate concentration has received little attention. The oversight is important.

JOHN D. BOWMAN, Ph.D. West Virginia 1972. Trends in the location of glass production.

EDWARD A. BRANCH, D.B.A. George Washington 1972. A study of selected procurement practices of small firms in metropolitan Washington, D.C. engaged in federally financed research and development.

This is a study of small Washington-based R&D firms that are currently engaged in federally financed R&D. The study develops a profile of selected organization, manpower, and marketing characteristics of the firms; examines the small business-government contracting process in terms of types of contracts used and preferred and methods of contract placement used and preferred; and examines the attitudes of Washington-based firms toward government contracting.

KEITH J. BREWER, Ph.D. McGill 1973. The Canadian east coast groundfish industry.

DAVID C. BUNTING, Ph.D. Oregon 1972. Rise of large American corporations: 1896–1905.

FOSTER BURTON, Ph.D. Pittsburgh 1972. A regional economic model of the construction industry.

This study relates construction in place in a time period to craft labor resources and the optimal mix of stocks of the several types of structures in a region by the use of a linear programming model. The importance of craft habits to the linear assumption is emphasized. Unemployment is examined and it is con-

cluded that some, previously assumed to be weather-related, is caused by shortages in a few crafts and is therefore structural.

BO A. W. CARLSSON, Ph.D. Stanford 1972. The measurement of efficiency in production: An application to Swedish manufacturing industries 1968.

The study attempts to bring together various definitions of efficiency in the economic literature; integrate inefficiency into standard microeconomic theory; measure efficiency empirically at the plant level in 26 Swedish manufacturing industries; and analyze the results in terms of macroeconomic variables. It is found that tariffs are negatively and the concentration ratio positively associated with efficiency, indicating that in a small, open economy the concentration ratio reflects economies of scale and specialization rather than market power.

GARY D. CARMAN, Ph.D. Southern Methodist 1973. Nuclear and gas-oil power plants as investment alternatives.

This study examines investment criteria by which regulated electric utilities choose new power plants. By means of sensitivity analysis, conditions are determined which make nuclear plants more economical than gas-oil plants.

MOHAMMAD A. CHAUDRY, Ph.D. Tufts 1972. The economics of competition in the U.S. communications industry: An econometric approach.

MICHAEL E. CONROY, Ph.D. Illinois (Urbana) 1972. Optimal regional industrial diversification: A portfolio analytic approach.

DANE J. COX, Ph.D. Cornell 1973. Toll road pricing: Theory and application.

ANTHONY W. CYRNAK, Ph.D. West Virginia 1972. The effect of regulation on external diseconomies: Surface mining in northern West Virginia, 1960–1970.

VICTORIA A. DAILEY, Ph.D. Virginia 1973. The certificate effect: The impact of federal entry controls on the growth of the motor common carrier firm.

This study sets forth a theoretical model of motor common carrier behavior with respect to investment in route authorizations. Under the present system of federal (ICC) controls of entry and route expansion, there will be a tendency for large motor common carriers to experience both greater growth and slightly lower costs than their smaller competitors. Econometric investigation of the two major hypotheses suggests that the model's implications were valid for 1958–68.

MICHAEL G. S. DENNY, Ph.D. California (Berkeley) 1972. Trade and the production sector: An exploration of models of multiproduct technologies.

LARRY J. DOBESH, Ph.D. Washington State 1973. Earnings control standards for regulated motor carriers.

This study examines the ICC standards for regulating the earnings of truckers. It determines whether the permitted earnings have been reasonable in terms of economic criteria. The ICC has employed irrational and biased standards to test the earnings of regulated truckers since the passage of the Motor Carrier Act of 1935. The operating-ratio standard provided no meaningful indication of the reasonableness of the earnings of regulated truckers. Deregulation of motor transport would offer the best solution for earnings control of trucking firms.

JEFFREY C. DONAHUE, Ph.D. Florida State 1972. The product and utilization alternative for bark in Florida and the costs of bark processing operations.

With few exceptions forest products firms in Florida traditionally have considered tree bark a waste wood-residue product generated under joint cost conditions with other marketable primary forest products such as sawlogs and pulp chips. The purpose of this study is to provide the comprehensive economic analysis which hopefully will enable forest products firms to achieve a greater, more proper utilization of their bark, and, concommitantly, enhance the development of the bark industry in Florida.

LOISJEAN DRAKE, Ph.D. Nebraska 1972. A statistical analysis of hospital costs in Nebraska.

This study provides evidence regarding economies (diseconomies) of scale in the general hospital industry using cross-section data for 68 community hospitals in Nebraska. Long-run average cost functions estimated for the fiscal years 1967–68 and 1968–69 exhibit behavior contrary to that predicted by economic theory. Diseconomies of scale predominate, with some tendency for average costs to decline after hospitals reach a size of about 330 to 400 beds.

DALE S. DRUM, Ph.D. Nebraska 1972. An economic analysis of the public power system in Nebraska.

Nebraska is the only state in the nation in which the entire electrical generation, transmission, and distribution system is publicly owned. Analysis reveals that plants are, in general, too small and rate policies ill-adapted to achieving high utilization. Thus, the comparatively low electrical rates in Nebraska are apparently attributable to the tax advantages of public ownership or to the failure of commissions in other states to exert effective earnings control.

GARETH P. DYAS, D.B.A. Harvard 1972. The strategy and structure of French industrial enterprise.

The thesis describes the evolution of strategy and structure, in France's 100 largest manufacturing companies, between 1950 and 1970. The main trends observed were towards increasing levels of product line diversity, and towards the eventual adoption of divisional forms of organization. They were explained in terms of economic, institutional, and cultural environmental factors. These trends have led to recent change

in the way of life within firms, with dramatic implications for French management.

WILLIAM R. EADINGTON, Ph.D. Claremont School 1973. The economics of gambling behavior.

Nevada is presently the only state with legalized casino gambling, although a number of other states are actively considering legalization. This study constructs a behavioral model of Nevada's tourist gambler and tests the model's implications against the four main resort areas. It is argued that the type of gambling motivation (entertainment or investment) depends mainly on income level. Predictions of the likelihood of success of casinos in other states are drawn from the model. Economic impact of gambling on Nevada is also discussed.

THOMAS W. EWART, Ph.D. Purdue 1972. Biddings for oil leases: A simulation approach.

MOHAMED A. FINAISH, Ph.D. Southern California 1972. Libya's petroleum policy.

It is argued that the typical oil producing country can be conceived as engaged simultaneously in a bargaining game with oil companies and a competitive game with other exporting countries. Within this framework, the Lybian government's petroleum policies since 1955 are analyzed and evaluated. Then, alternative strategies available for oil-exporting countries, chiefly increasing host country participation and nationalization, and their advantages and limitations are discussed with special reference to Lybia.

THOMAS S. FRIEDLAND, Ph.D. California (Berkeley) 1972. Advertising and concentration.

PETER V. GARROD, Ph.D. California (Berkeley) 1972. Service and employment implications of food distribution: A study of the retail sector of Santiago, Chile.

The dual role of the distribution sector of providing market services and employment opportunities is examined. The existing fragmented, small-scale structure characteristic of the low-income regions is compared with the more capital intensive mass-merchandising operation characteristic of the high-income regions. Welfare comparisons indicate that the gain in consumer surplus resulting from the lower prices charged by the large-scale operations would more than compensate the displaced employees of the existing small firms if these firms were to be replaced by the larger more labor efficient operations.

MALCOLM M. GETZ, Ph.D. Yale 1973. The incidence of rapid transit in Atlanta.

PHILLIP E. GIFFIN, Ph.D. Tennessee 1972. Industrial concentration and firm diversification in bituminous coal with special reference to the southeastern United States, 1950–1970.

This study presents an analysis of concentration and changes in concentration in bituminous coal. The approach is to view concentration in various markets with major emphasis on a southeastern submarket. Concentration figures are made available for buyers and sellers. The study also presents data on coal firm diversification. Essentially, the findings indicate rising concentration levels in addition to increasing diversification. Another contribution of the study is the method and data supplied in defining a submarket.

PAUL R. GILON, Ph.D. Southern California 1972. A constrained price discriminator: An application to the U.S. Post Office.

In this study the behavior of a profit maximizing third degree price discriminator operating in two markets is analyzed. The firm is further subject to the constraint $P_1 = (1\text{-}d)P_2$, where d is a policy set discount rate and P_1 and P_2 are the prices. Specifically, the impact of the discount rate d is examined under various assumptions. Subsequently, as an application, the theoretical models are fitted with 30 years of U.S. Post Office data. Finally, various conclusions with their respective assumptions are presented and discussed.

ROBERT S. GOLDEN, Ph.D. California (Berkeley) 1972. Optimal hedging decisions for beef cattle feeders.

The study is concerned with the development of a cattle purchase-hedging decision model. The model draws on the mean-variance approach to asset selection and Holbrook Working's theory of hedging storable commodities. Price expectations are required for the model and are formalized through the use of an econometric forecasting system. Spectral analysis is used in the identification and estimation of the system. The study is concluded through an example of the use of the model.

ALICE W. C. GORLIN, Ph.D. Michigan 1973. Soviet firms and the rationalization of the shoe industry of the *USSR*.

JAMES W. GOTCHER, JR., Ph.D. California (Berkeley) 1972. A systems approach to policies governing the design of an information dissemination system for the earth resources technology satellites.

ROBERT T. GREUBEL, Ph.D. Arkansas 1973. An evaluation on the performance of the electrical machinery industry: A case study of an industry.

The null hypothesis is that the action taken on the part of the government in the electrical conspiracy case did not make any difference in the performance of the transformer industry. In many ways the preconsent and postconsent time periods exhibit no difference in the performance variables analyzed. The possibility that collusive and noncollusive oligopoly may produce the same results and the data used may render this conclusion an over-simplification.

ROBERT HALVORSEN, Ph.D. Harvard 1973. Residential demand for electricity.

Long-run elasticities of demand are estimated with pooled cross-section data. Because price is a function

of quantity purchased, a simultaneous equations model is estimated by *TSLS*. Estimates of long-run elasticities are robust. Electricity price is the most important determinant of demand. Short-run elasticities are estimated for individual states with time-series data. The short-run estimates are sensitive to the choice of dynamic model.

HOMER J. HOLLAND, D.B.A. George Washington 1972. On the relative accuracy of the Delphi method as an aid in technological forecasting.

An experimental study examining the relative accuracy of three methods: Delphi, Committee, and Questionnaire, for eliciting responses from groups of experts. The study concludes that groups produce more accurate estimates when they use the Delphi Method as compared to the Committee or Questionnaire Methods.

WASFY B. ISKANDER, Ph.D Indiana 1973. The economic potential of copper in the Arctic.

This study presents an analysis of the economic potential of copper in the Canadian Arctic, also referred to as the North, during the 1973–1985 period. The study deals specifically with the Yukon and Northwest territories. The primary purpose of this study is to provide guidelines for appropriate amounts and direction of new investments on infrastructure for copper production.

MD NURAL ISLAM, Ph.D. McGill 1973. Interregional competition and comparative location advantages of North American newsprint industry.

GLENN P. JENKINS, Ph.D. Chicago 1972. Analysis of rates of return from capital in Canada.

The principal objective of this study is to estimate the rates of social and private return and effective taxation that are produced by physical capital in Canada. This information is developed on a detailed sectorial breakdown for the years 1953 to 1969, and is used to evaluate the taxation gains from foreign owned capital, the social opportunity cost of public funds, and the economic cost of distortions created by the nonneutral taxation of income from capital.

PHILIP J. KARST, Ph.D. Washington (St. Louis) 1972. The effect on performance of the separation of ownership from control.

PETER KEMPER, Ph.D. Yale 1973. The location decisions of manufacturing firms within the New York metropolitan area.

MANFRED KIMMLE, D.B.A. Harvard 1972. Design of management information systems: An approach and case study.

This thesis examines how one should determine the content and structure of a management information system (*MIS*) to support executive decision makers in managing the overall business. It presents a normative framework for *MIS* design, demonstrates its application at a large hospital organization, and provides an authoritative appraisal by a panel of experts. As precedents rather than prescriptions, the propositions and findings should point the way to more effective *MIS* elsewhere—in the health care sector, in government, and in private industry.

GEOFFREY R. KING, Ph.D. Southern California 1972. Arbitration of technological change.

This study investigated the reported industrial labor arbitration cases involving technological change between 1960 and 1970. The cases were analyzed for evidence of guiding principles and a common thread of rulings. The analysis covered union and company contentions, and the results of those contentions. There were eight conclusions from the study.

CAROLE E. KITTI, Ph.D. Chicago 1973. Patent policy and the optimal timing of innovations.

REUBEN KYLE III, Ph.D. Tennessee 1972. Seller concentration, price competition, and nonprice forms of competition in consumer goods industries.

A number of theoretical analyses are consistent with the hypothesis that sellers tend to substitute nonprice competition as seller concentration increases. A test of the hypothesis was devised using price flexibility, defined in various ways, to represent the extent of price competition and advertising as a proxy for all forms of nonprice competition. A sample of forty consumer goods industries offered support for the hypothesis.

JULIAN E. LANGE, Ph.D. Harvard 1973. The bidding process in the construction industry.

A general framework for the analysis of the bidding process in construction is developed. The framework is then applied to the bidding process in a specific industry submarket: i.e., state-initiated building construction in Massachusetts. Multiple regression techniques are employed to explain variations in contractor bids. The results suggest that the approach is a fruitful one and indicate certain promising avenues for extensions of the research.

DAVID P. LEVINE, Ph.D. Yale 1973. Accumulation and technical change in Marxian economics.

EDWARD B. LEVITON, Ph.D. State University of New York (Binghamton) 1973. A cash flow theory of conglomerate mergers.

This paper explores the hypothesis that with passage of the Celler Kefauver Amendment (1950), effectively eliminating vertical and horizontal merger opportunities, the conglomerate form of merger has developed as an attempt to maximize growth consistent with profit maximization. Results of multivariate analysis appear to support the assertion that cash flow generation did play significant role in explaining merger activity during 1960's.

CHANTEE LEWIS, D.B.A. George Washington 1972.

Simultaneous equation production functions (Cobb-Douglas type) for decisions pertaining to seabased tactical air.

The need for analytical aids to assist management in making input allocation decisions concerning the operating of military aircraft is quite clear. A number of authors have suggested that a Cobb-Douglas type production function exists in this situation. An exponential model is postulated and five aircraft cases are analyzed. R^2 of .884 to .951 were obtained. The paper concludes that an exponential model is appropriate for military aircraft capital versus labor investment decision.

CHARLES A. LUCKETT, Ph.D. Georgtown 1973. A model of the markets for automobile sales and credit.

Three alternative models are developed to explain quarterly automobile unit sales and credit, 1961–70. The best performing model contains four equations, solved recursively, for yearly changes in auto sales, credit extended, the Commerce Department's Buying Intentions Index, and a loan terms index based on contract maturities and down-payment percentages. Loan terms prove a significant determinant of credit usage. Buying intentions and employment variables are found superior to income variables in explaining sales.

RUFUS A. LYMAN, Ph.D. Northwestern 1972. Demand conditions in the electrical power industry: An econometric study.

The dissertation investigates demand conditions in the residential, commercial, industrial, and public street and highway lighting categories of customers. Data are for utilities and were collected for the period 1959–68. The model is specified with a general functional form of which the linear and log-linear forms are special cases; non-linear regression methods are used. Results of interest concern the specification of interaction terms, climatic variables, and income distribution variables. Also of interest are price elasticities and income redistribution effects.

THOMAS W. MASON, Ph.D. Pittsburgh 1972. The effects of entry: A study of the American aluminum industry.

This thesis examines the effects of the structural changes due to the entry of new firms into primary aluminum production in the United States. Alternative models of price determination are formulated, and their predictions are compared with the actual prices for primary aluminum ingot for the years 1923–40 and 1947–68. Results indicate that the entry of additional manufacturers brought about changes in conduct and lower prices for virgin aluminum in the postwar market.

JOHN C. MOORE, JR., Ph.D. California (Berkeley) 1972. The least cost organization of cotton ginning facilities in California's San Joaquin Valley.

The basic objective of the study was to determine the size number, and location of cotton gins in the San Joaquin Valley that would minimize assembly and processing costs. Costs were developed through the cost synthesis approach and the final solution was derived using a variation of the linear programming transportation model. The analysis included extended ginning seasons and seed cotton storage. It was found that San Joaquin Valley cotton farmers could reduce costs considerably if the least cost system were adopted.

FREDERICK J. E. MORRISSEY, D.B.A. George Washington 1972. The law and licensed lending: Georgia in the 1950's.

The twin objectives were to portray the personal loans market served by the consumer finance industry and to determine the bases for the high ranking in loan service of the industry in Georgia. The market in Georgia was found to be not one market, but to comprise three distinct, noncompetitive markets. The high degree of loan services were ascribed to: the historical evolution of consumer finance in the state; the present law which provides for borrower's protection, handsome charges, and generous loan ceiling; and the liberal interpretation of the convenience and advantage clause.

RICHARD D. MURPHY, Ph.D. Michigan 1972. The determinants of the extent of multiplant operation in U.S. manufacturing industries.

ROBERT H. NICHOLSON, Ph.D. North Carolina State 1972. An economic evaluation of alternative food service systems for the North Carolina Department of Correction.

This study examines the relative costs of alternatives to the food delivery system operated by the North Carolina Department of Correction. After describing economic relationships among expenditure categories, and identifying technologically feasible systems, partial budgeting techniques are used to select the food delivery system offering maximum budgetary savings. A centralized production and distribution system operated by Department of Correction personnel was found to give the lowest projected budget. The future possibilities of wage payments for inmate labor and use of noninmate employees in the central commissary do not change this conclusion.

YUKIO NOGUCHI, Ph.D. Yale 1972. An economic theory of diffusion of knowledge.

YEUH-YUN CHENG O'BRIEN, Ph.D. Illinois (Urbana) 1973. Intertemporal equilibrium models of a monopoly and the analysis of tax effects.

GERALD W. OLSON, Ph.D. Rutgers 1973. Consumer demand for services in the United States, 1947–69.

Five dynamic models of consumer demand for services are tested to determine the causes for the rapid growth of service expenditures and service prices since 1947. Using the results of these estimates, expenditure

projections and policy simulations are made to determine the optimum set of policy actions to stem the inflationary pressure arising from the service sector.

FRANKLIN L. ORTH, JR., Ph.D. Tennessee 1970. An empirical analysis of the relationship between diversification and profitability in the 1000 largest industrial corporations, 1965.

HOWARD E. OVERCAST, Ph.D. Virginia Polytechnic Institute 1973. The survivor technique: Reappraisal and revival.

The survivor technique here uses a statistically consistent chi-square test to compare the temporal distribution of plants. Minimum efficient size is estimated by a statistical rule which is based on a hypothesis concerning the relevant distribution. This rule yields results somewhat similar to those obtained previously. The particular advantages of this particular appraisal are: a longer time period, a larger sample, and the use of a consistent statistical rule rather than a rule-of-thumb.

UN-SUH PARK, Ph.D. Pittsburgh 1972. A study of technical progress in the U.S. aggregate economy.

The biased disembodied technical progress under neoclassical production function model is inconsistent with the near constancy of the share ratio if strict constancy of the elasticity of substitution is adhered to. Certain law of movement of elasticity of substitution is established. The varying elasticity of substitution is also estimated by capturing the shifting characteristics of *CES* production functions in a way which is statistically consistent. This also leads to the estimation of biased technical progress.

ROBERT A. PITTS, D.B.A. Harvard 1972. Variable incentive compensation for division general managers in diversified companies.

This thesis compares variable incentive compensation schemes for division general managers in eleven large diversified firms, six of which have grown primarily through internal development (Borg-Warner, Corning Glass Works, General Electric, Honeywell, 3M, and Westinghouse), five primarily through acquisition of other companies (AMF, Gulf & Western Industries, ITT, Midland-Ross, and Sybron). Incentive compensation awards for division general managers are found to be generally more affected by corporate performance, subject to more top management discretionary influence, larger, and subject to a wider payoff range in the former than the latter group of firms.

RAWLEIGH H. RALLS III, D.B.A. George Washington 1971. A time-series analysis of the nature of the demand for life insurance.

This study investigates at the industry level the relationships that exist between aggregate consumption and savings and the sale of life insurance products. Investigation at the company level is also carried out with data of considerably more detail than that available at the national level. The investigations reveal that protection appears to be purchased when savings are high, and the savings component of a life insurance product appears to be purchased when consumption is rising.

FREDERIC D. RANDALL, D.B.A. Harvard 1972. Corporate strategies in the drug industry: A study of strategic response to social and political pressures.

This thesis examines strategic response within firms in one industry that has experienced significant social and political pressures. It describes what strategies emerged and why. Field research was conducted in six drug firms to determine strategic response in several areas of activity. Strategies fell into one of three patterns depending on each firm's internal characteristics in two key areas—innovation and vulnerability. These dictated strategies in product, marketing, pricing, management, finance, and in consumer and government relations.

ANTHONY A. ROMEO, Ph.D. Pennsylvania 1973. Interindustry differences in the diffusion of an innovation.

JOSEPH M. ROOP, Ph.D. Washington State 1973. The economic organization of state agricultural experiment stations.

The behavior of a specific type of nonprofit organization, state agricultural experiment stations, is examined. The analytical tool is derived from that branch of the theory of organizations most closely associated with the names of Chester I. Barnard and Herbert A. Simon. The inducements-contributions approach to organizational theory is used to develop a behavioral and structural model of experiment station activity. The model is capable of generating independent evidence bearing on a number of the suggested hypotheses.

HARVEY ROSENBLUM, Ph.D. California (Santa Barbara) 1972. The structural determinants of merger activity: 1948–1968.

GEORGE R. SANDERSON, Ph.D. Stanford 1972. The differential impact of macroeconomic conditions on industry profits.

JULES J. SCHWARTZ, D.B.A. Harvard 1973. The decision to innovate.

Exploratory field studies in two high-technology firms confirmed the applicability of Bower's resource allocation model to technological innovation decisions. Managers contributed to problem solving by articulating strategy, negotiating criteria and preparing supporting business plans. They managed risk by factoring problems and imposing schedules in return for their support. Large rewards to innovators were legitimized by sharing responsibility for outcomes. Proposals assumed a conservative bias and were incremental rather than radical in nature.

ELLIOTT D. SCLAR, Ph.D. Tufts 1972. The public

financing of transportation in the greater Boston area.

JOHN J. SIEGFRIED, Ph.D. Wisconsin (Madison) 1972. The relationship between economic structure and the effect of political influence: Empirical evidence from the Federal Corporation Income Tax Program.

This study attempts to demonstrate that systematic empirical analysis is possible for the important issues overlapping the area of political science and economics. The sample consists of 1963 corporation income tax information and economic variables aggregated by IRS "minor industry." In general, it is found that firm size is negatively correlated with the effective average corporation income tax rate; market concentration and the profit rate are positively related to the tax rate; and industry size and the geographical dispersion of an industry's employment are not associated with variations in the tax rate.

JAMES H. SOOD, D.B.A. George Washington 1972. A study of the international computer industry and a projection of the effect of the American computer industry on the U.S. balance of payments from 1969 to 1975.

It is anticipated that the American computer manufacturing industry will have an increasingly larger impact on the *U.S.* balance of payments in the future. This study is concerned with the development of the projection of the effect of this industry on the *U.S.* balance of payments from 1969 to 1975. This projection should then provide a more complete understanding of the economic and financial strength of the United States during this period.

JOHN R. STODDEN, Ph.D. Illinois (Urbana) 1972. Firm size, industry structure, and cost of capital.

HORST SYLVESTER, Ph.D. Michigan 1972. New and used cars as consumer alternatives.

RICHARD F. SYRON, Ph.D. Tufts 1972. An analysis of the collapse of the normal market for fire insurance in substandard urban core areas.

DAVID J. THOMSEN, Ph.D. Oregon 1973. The analysis of conglomerate performance.

DAVID TIERNEY, Ph.D. Temple 1972. Employment effects of textile imports.

THOMAS F. TORRIES, Ph.D. Pennsylvania State 1972. Economic descriptions and simulation of the construction aggregate industry in the Allentown-Bethlehem-Easton area of Pennsylvania.

The industrial and chemical stone industry is described in economic terms. The demand for industrial and chemical stone is forecast for the Allentown, Bethlehem, Easton, Pennsylvania area for the years 1968 to 1975. The potential study of raw materials to satisfy the demand for aggregate materials is identified in the study area. A computer simulation model is constructed and operated to select the most profitable supply areas. An optimal solution occurs when the overall gross cost to the consumer is minimized and the profits of all the individual producers are maximized, subject to all producers receiving a specified minimum rate of return on sales and subject to all demand for all materials being satisfied.

JOHN M. TRAPANI, Ph.D. Tulane 1972. The role of money in the theory of production: A theoretical and empirical analysis.

The purpose of this thesis is to develop and test a microeconomic model that explains the sources of productivity of money balances held by firms. It is shown conceptually that money balances conserve on entrepreneurial effort in acquiring conventional inputs, and that such balances may reduce unit production costs. This implication is tested using cross-section samples of electric utilities and the above hypothesis cannot be rejected. The influence of monetary intensity upon the firm's equilibrium capital-labor ratio is also analyzed theoretically and empirically.

ANTONY B. T. WERNER, Ph.D. Pennsylvania State 1973. The financing of the Canadian metal mining industry, 1945–1960.

The demand for and supply of funds to the Canadian metal mining industry in the 1945 to 1960 period is examined. Foreign demand for Canadian mineral products is considered to have provided the chief stimulus for the expansion of the Canadian metal mining industry. Techniques used to finance the mining industry are outlined. Returns to net capital are found to have declined, as did returns to investors between 1954–59 and 1960–65. No definable relationship is found to exist between risk and reward in the case of the Canadian mining industry.

WILLIAM W. WILES, Ph.D. Wisconsin (Madison) 1973. Identifying characteristics of acquired, acquiring, and nonmerging banks.

ALBERT L. WINKLER, Ph.D. Kansas State 1972. Investment among the states.

DARYL N. WINN, Ph.D. Michigan 1973. The relations among industrial profitability measures, concentration, firm size, capital intensity, growth, and business risk, 1960–1968.

PAUL B. ZIMMER, Ph.D. California (Los Angeles) 1973. Benefits and costs of peak period congestion tolls on the Tokyo urban expressway network.

Agriculture and Natural Resources

ROBERT L. ADAMS, Ph.D. Illinois (Urbana) 1972. An economic analysis of the junk auto problem.

ANTONIO AGUIRRE, Ph.D. California (Berkeley) 1972.

Welfare cost of protection: The fertilizer industry in Argentina.

The object of this study is to analyze the situation of the nitrogenous fertilizers industry in Argentina, which is an example of the implementation of the industrialization policy via import-substitution. An attempt is made at estimating the lower limit of the welfare cost of protection of this "infant" domestic industry. Finally, an evaluation of the social costs and benefits of alternative policies is made and its implications analyzed.

MUDDATHIR A. AHMED, Ph.D. Arizona 1972. Electronic data processing as a tool in farm management: A case study of Arizona AMAP users.

The bulk of electronic data processing programs now in use by both public and private agricultural concerns are directed at the following three objectives: to supply records for tax and financial purposes; to supply cost and return analysis for enterprises and general farm management; and to provide a source of data for research and educational purposes. However, actual use by farmers is concentrated on the first objective.

DOUGLAS R. A. BECK, Ph.D. Iowa State 1972. Water quality and community viability.

This project determines implications for regional growth when water quality parameters were imposed on specific industries. A quantitative model was developed as the basis for computer simulation of an eight-county Iowa F.E.A. to provide information and projections for use in planning service and industrial development under water quality constraints. An environmental impact program permits an analysis of major demographic-economic interactions, their relationship to water quality controls and their impact on specific sectoral reallocations of employment, output, and investment within the regional area.

FRED J. BENSON, Ph.D. Missouri 1972. Technical production function estimates for corn which include weather variables.

JACQUES BLAS, Ph.D. Missouri 1972. Economic analysis of the influence of resources restrictions on farm output and total income in a region of Spain.

ALDO L. BIONDOLILLO, Ph.D. Minnesota 1972. Social cost of production instability in the grape-wine industry: Argentina.

MICHAEL F. BRADLEY, Ph.D. Cornell 1973. Proposal for long-term apple processing marketing contracts.

This study analyzes traditional grower-processor relationships in the apple industry and the problems associated with price making. Included are price and cost differentials for specific quality variations which form the basis for long-term forward formula pricing. The study concludes with model contracts including a number of alternatives to satisfy individual grower-processor needs.

MAURICE H. BRUCKER, Ph.D. Illinois (Urbana) 1973. Computer-based retirement planning system for farmers with low income.

CLARK R. BURBEE, Ph.D. Minnesota 1972. Adjustments to improve the competitive position of the upper midwest turkey industry.

BRUCE O. BURNHAM, Ph.D. Colorado State 1972. The economics of technical change: With special emphasis on Colorado agriculture.

Development and application of methodology which quantifies the impact of resource development and technological change on gross output in the agricultural sector of a multi-county region has lagged the specification of similar models at the national level. In an attempt to narrow the gap, this thesis develops economic models which relate gross output to intermediate inputs and technical change inputs for the agricultural sector of Colorado and multi-county regions of the state.

GERALD R. CAMPBELL, Ph.D. Michigan State 1973. Procurement contracting and sow leasing as vertical coordination arrangements in the hog-pork subsector.

JOSÉ L. CARVALHO, Ph.D. Chicago 1972. Production, investment and expectations: A study of the United States cattle industry.

Conciliation between production and investment theories is attempted through the use of a dynamic model which implies profit maximization over time. The maximization problem is solved with the help of an IBM/360 by means of a computer language, FORMAC, which allows algebraic manipulations. Expectations formation about future prices prove to be very important for the profit maximization, and a "quasi-rational" model based on time-series analysis is proposed. Empirical tests are performed with satisfactory results.

DONALD G. CHAFIN, Ph.D. Ohio State 1973. Economic analysis of futuristic beef cattle and forage production systems.

SEAN B. CHIN, Ph.D. Missouri 1972. Exhaustible resources and technological change in U.S. agriculture, 1940–1969.

JON M. CONRAD, Ph.D. Wisconsin (Madison) 1973. Uncertain externality: The case of oil pollution.

Many substances which generate external diseconomies also involve elements of uncertainty; specifically, the extent of the divergence between social and private cost may not be known in the present. This phenomenon was entitled uncertain externality and was seen to arise from random elements of certain synthetic and natural systems. Oil pollution is a special case of uncertain externality. The uncertainty arises from: random elements of the (synthetic) closed subsystem; unpredictable physical conditions present

when a spill occurs; and incomplete biological information on the effects of crude and refined oils on marine populations.

SANDY A. D'AQUINO, Ph.D. Colorado State 1972. Optimum allocation of natural resources.

This study develops a general formulation of a deterministic linear programming model that can be applied to management of resource systems. This deterministic model would, in turn, be related to a stochastic model which would determine discount rates for "risk" in coefficients of the resource model. The study had three main objectives: to assist managers in optimum allocation of resources; to adjust allocation of resources as a result of variance-covariance in coefficients of the model; and, to calculate necessary adjustment in the discount rate as a result of variance-covariance in product prices.

LEROY DAVIS, Ph.D. Illinois (Urbana) 1972. Relationship of market structure to spatial and temporal pricing patterns in the grain industry.

ALVIN J. DE BOER, Ph.D. Minnesota 1972. Technical and economic constraints on bovine production in three villages in Thailand.

Thai bovine production has lagged behind the rest of agriculture. A one year study of three Thai villages was used to model village bovine resource allocation. Marginal value products of crop production resources were estimated and charged against bovine enterprises. Internal rates of return were estimated for ten types of production represented. These indicated few inefficiencies in the existing system. Further models were constructed incorporating technical change and its likely effects on the system.

RALPH D. ELLIOTT, Ph.D. North Carolina State 1972. The industrial demand for water and waste treatment in selected U.S. cities which are levying surcharges.

A study of thirty-three cities which have been levying surcharges based on the strength of industrial wastes showed that these charges provide a worthwhile incentive for firms to recover or otherwise dispose of wastes that are now flushed away with water. Industrial demands for water and waste treatment are both negatively related to marginal charges for water and surcharges, as expected. Elasticities are thought to fall in the range of —0.5 to 1.0.

THEODORE J. ELLIS, Ph.D. Colorado State 1972. Role of oil shale in U.S. energy mix.

ELIZABETH B. ERICKSON, Ph.D. Illinois (Urbana) 1972. A linear programming model for analysis of irrigation agriculture in northwest Mexico: A pilot study of three irrigation districts.

MARVIN R. FAUSETT, Ph.D. Missouri 1972. An empirical study of the economic growth and capital formation on selected Missouri dairy farms.

RICHARD S. FENWICK, JR., Ph.D. Missouri 1972. Capital acquisition-strategy for Missouri farm supply cooperatives.

JULIA M. FRIEDMAN, Ph.D. Oregon 1973. Minimizing the cost of water quality control for the Willamette River Basin.

NORBERTO FRIGERIO, Ph.D. Michigan State 1973. Alternative wholesale facility arrangements for fresh fruits and vegetables in the Buenos Aires metropolitan region.

JOHN C. GAMBLE, Ph.D. Illinois (Urbana) 1973. Economic evaluation of the impact of alternative environmental regulations and waste management systems on Illinois hog producers.

GERALD L. GEHRE, Ph.D. Arkansas 1973. The measurement and role of dairy producer's economic perception and attitudes toward class I base plans.

The primary objectives of this study were to measure dairymen's economic perception and attitudes and to relate these attributes along with socioeconomic variables to preferences for base plans. Some of the statistical tools employed were Mann-Whitney U-test, chi-square and multiple regression. Conclusions were that policy makers should consider: regional variances in perception and attitudes; the "free rider" problem in a voluntary base plan; and attitudes and socioeconomic factors in developing an economic program.

ADAM GIFFORD, JR., Ph.D. California (San Diego) 1972. Environmental issues and economic growth.

Pollution, natural resource constraints, and recycling are considered within the context of aggregate growth models. The analysis describes the types of reactions the economy can make when confronted with such environmental problems. The results suggest the possibility of maintaining a viable environment along with economic growth.

GURMUKH S. GILL, Ph.D. California (Berkeley) 1972. Demand and supply of wastepaper with policy implications for quality of the environment.

Employing available empirical evidence, a case is made for increased recycling of wastepaper in urban refuse. Despite this, increased recycling is considered unlikely until economic constraints on the reuse of wastepaper are recognized and removed. To help identify such measures, the demand and supply functions for wastepaper in the United States are econometrically estimated and the computed elasticities are used as measures of relative effectiveness of various constraints. Some measures to reverse or offset the identified constraints are recommended while others are suggested for further investigation.

DARREL L. GOOD, Ph.D. Michigan State 1972. Potential impact of environmental pollution abatement alternatives on the Michigan dairy farming industry.

RICHARD D. GREEN, Ph.D. Missouri 1972. Expectation formulations and optimal decisions in cattle feedlot problems.

ELLIOT S. GROSSMAN, Ph.D Rutgers 1973. National and state level econometric models of the U.S. asparagus industry, 1948–1969.

An examination of static supply and demand models on the national and state levels including farm level relationships for all asparagus, and for fresh and process asparagus taken separately. Included are forecasts for 1970 and 1980. A production function is estimated using the pooling method. A geometric distributed lag model of the national all-asparagus supply is estimated via a modification of Wallis' method and by Dhrymes' double search method.

JORGE GUTIERREZ-VILLARREAL, Ph.D. North Carolina State 1972. Investment alternatives in the processing of North Carolina apples.

The profitability of a variety of apple processing activities was evaluated for three sizes of plants operating for three lengths of season. Initial investment, equipment specifications, and operating costs were calculated for single and multiple-product plants which produced canned slices, sauce, frozen slices, juice, vinegar, concentrate and essence and apple butter. Calculated internal rates of return indicate that juice and juice-sauce in combination have the greatest profit potential. The effects of level of raw and finished product prices, conversion rates and raw product availability on rates of return are examined.

TEKETEL HAILE-MARIAM, Ph.D. Stanford 1973. The production, marketing, and economic impact of coffee in Ethiopia.

Coffee accounts for more than 50 percent of Ethiopia's export earnings and nearly 13 percent of its monetary *GDP*. This analysis measures the coffee industry's gross value-added and net contributions to national income, contributions of foreign exchange and savings to noncoffee sectors, and linkages with the domestic economy. Findings are that the popular view that Ethiopian coffee comes from natural forests is incorrect; the supply of coffee is low relative to domestic and export demand, and export demand and expansion of exports have not been seriously constrained by the world quota system.

OSMAN A. HAKIM, Ph.D. Arizona 1972. The effects of the U.S. cotton policy on the world market of extra-long staple cotton.

The Sudan of which the author is a citizen is one of the major producers of extra-long staple cotton. On the other hand, the United States is the world's largest producers of cotton. The *U.S.* cotton policy is assumed to have some effects on the world market for extra-long staple cotton.

RONALD J. HANSON, Ph.D. Illinois (Urbana) 1972. An economic analysis of off-farm income as a factor in the improvement of the low farm income farmers in Illinois.

ALLAN J. HARRIS, Ph.D. Kansas State 1973. Economic incentives in dryland crop-share leases, western Kansas, 1951 and 1970.

WAYNE HAYENGA, Ph.D. Michigan State 1973. The effects of bank mergers on financial services available to rural Michigan residents.

ROBERT B. HELMS, Ph.D. California (Los Angeles) 1973. The effectiveness of Federal Power Commission regulation of the field market for natural gas.

The effectiveness of the FPC's regulation of the natural gas field market is tested by estimating supply relationships for the period 1945–60 and then comparing projected and actual prices and reserves for the period 1961–69. The study concludes that area rate regulation caused a shortage of gas only after 1967. The control of field prices is shown to have provided an monopsony benefit to regulated pipelines at the expense of natural gas producers and consumers.

RICHARD C. HOYT, Ph.D. Minnestoa 1972. A dynamic econometric model of the milling and baking industries.

This dissertation encompasses variables such as the *U.S.* demand for baked products, domestic and export demand for wheat, *U.S.* and foreign wheat production capabilities, domestic and overseas demand for hard red winter and spring wheats, and the domestic demand for millfeed in a simulation framework in order to determine, for example, the impact of different hypothetical government programs on crop prices and prices and profits in the milling and baking industries. The model is also amenable to experimentation with respect to any of the other approximately 100 variables contained therein.

JAMES D. HOWELL, Ph.D. Ohio State 1973. Large farm organization in east central Ohio.

ROBERT D. HUNT, Ph.D. Minnestoa 1972. The contrasted effects of quota, autarky and free trade policies on U.S. beef production and prices.

The study attempts to estimate the effects of the U.S. Meat Import Quota on *U.S.* beef prices and production and make some assessment of the resultant welfare costs. A block recursive system consisting of a single equation beef import sector and a simultaneous equation domestic beef production and consumption sector was estimated. Analysis showed welfare costs to be significant though not large. The income transfer effect was also estimated and related to the social cost of the transfer mechanism.

FRANCIS S. IDACHABA, Ph.D. Michigan State 1972. Marketing board taxation of Nigerian agriculture: Some implications.

BOUWO M. JELLEMA, Ph.D. North Carolina State

1972. Analysis of the world market for groundnuts and groundnut products.

Using a market simulation model, which included twenty-five country groups and four seasons of the year, the most profitable markets and seasonal trade flows were identified for all exporters of groundnuts under free trade conditions taking into account costs of transportation and storage. The effects of various tariff policies of the *EEC* and the United Kingdom were calculated in terms of changes in producer and consumer surpluses and in government revenues. The overall net effects of the tariffs were found negligible, but the income transfers substantial.

FARRELL E. JENSEN, Ph.D. Kansas State 1972. A comparison of demand and factors affecting consumer acceptance of fresh and frozen meat as determined from a retail market test.

RICHARD E. JUST, Ph.D. California (Berkeley) 1972. Econometric analysis of production decisions with government intervention: The case of the California field crops.

Decision theory and the economic theory of risk-bearing are used to develop an econometric model of decisions made with changing risk. The effects of government farm programs are then included. The model is used to investigate the effects of government programs on acreage decisions in California. Acreages regulated directly by marketing quotas appear to be highly influenced by their programs while crops with voluntary programs are influenced more by risk and quotas for competing crops.

JAMES B. KLIEBENSTEIN, Ph.D. Illinois (Urbana) 1972. An economic analysis of the effects specialty grains should have on production decisions made by farmers.

PHILLIP L. KNOX, Ph.D. California (Berkeley) 1973. The U.S. beef cattle industry: A policy analysis with costs of adjustment theory.

The three segments of the study are the development of a theoretical model including costs of adjustment for input changes, the estimation of the model, and the application of the model to the evaluation of current beef import policies. In the imperfect, dynamic setting with largely naive expectations, producer incomes were found to decrease as a result of the beef quota on importation. Slow adjustment negated immediate income gains and overadjustment for the longer period resulted in lower prices and income.

JERRY J. KROPF, Ph.D. Missouri 1972. Social indicators of efficiency and equity convergence in four distressed regions.

HYDER A. G. LAKHANI, Ph.D. Maryland 1972. Some static and dynamic relationships between economic development and environmental quality.

This study attempts to find the nature and extent of relationship between economic development and environmental quality. The static framework hypothesizes an inverse relationship between economic development and environmental quality by means of transformation curves and social indifference curves and tests it by multiple regression equations for air quality in 100 *SMSA*s. The dynamics postulates a Gompertz function for diffusion of environment-saving techniques in petroleum refining and tests it by two-stage least squares.

LEELAND D. LAMBERT, Ph.D. Michigan State 1973. Regional trends in the productivity of American agriculture.

KYUNG W. LEE, Ph.D. Illinois (Urbana) 1973. Multivariate analysis and geographical comparison of the demand for grain dryers in four midwestern states.

DAVID A. LINS, Ph.D. Illinois (Urbana) 1972. An analysis of sources and uses of funds in the farm sector of the United States.

KENNETH J. MCKENZIE, Ph.D. Michigan State 1972. An economic analysis of selected agricultural extension programs in Colombia, South America.

NIKOLAOS S. MARTINOS, Ph.D. Iowa State 1973. a) Econometric models of the labor market in the farm sector of the north central region of the U.S.A.; b) The demand for farm labor in the U.S.A.

HARVEY A. MEIER, Ph.D. Ohio State 1972. Agricultural lending attitudes, practices, and activities of Ohio commercial bankers.

CEVDET OGUT, Ph.D. Missouri 1972. Technical response functions for corn—an economic analysis of nitrogen fertilizer experiments.

KAO-TWAN ONG, Ph.D. Illinois (Urbana) 1973. Quantitative analysis of applying random coefficient regression model to the short-run fluctuations for hogs.

HARUO ONISHI, Ph.D. Illinois (Urbana) 1973. Spatial and temporal resource allocation methods for agricultural watershed planning.

INJA K. PAIK, Ph.D. Georgetown 1972. Measurement of environmental externality in particular reference to noise.

Beginning with an exposition of the general concept of externality as related to market failure and the divergence of social and private costs, and the environmental pollution as a case of externality, the dissertation develops a method for measuring environmental externality through testing the hypothesis that the external diseconomy of aircraft noise is capitalized negatively into residential property values. If two properties are identical except being subjected to two different levels of noise, the difference in the price repre-

sents the *perceived* damages attributable to the difference in the noise.

ALFRED L. PARKS, Ph.D. Illinois (Urbana) 1973. The effects of wage contract provisions in the contracts of wholesale milk drivers on total wage payments and unit labor delivery costs in the fluid milk industry.

DAVID L. PEACOCK, Ph.D. Michigan State 1972. The adoption of new agricultural practices in northeast Brazil: An examination of farmer decision making.

STEVEN M. PEASE, Ph.D. Yale 1973. The spatial agricultural economy: A theoretical study with special reference to Brazil.

FREDERICK M. PETERSON, Ph.D. Princeton 1972. The theory of exhaustible natural resources: A classical variational approach.

A partial equilibrium, present value maximizing model of an extractive industry is solved by classical variational methods with the assumption of continuous time, perfect certainty, and an infinite time horizon. The basic model is augmented to include technical progress, exploration, and recycling. The effects of monopoly, externalities, and tax policies on the time paths of exploration, mining, recycling, and prices are determined by a priori reasoning and by numerical analysis.

DAVID C. PETRITZ, Ph.D. Illinois (Urbana) 1972. Economic implications of the effects of climatic stress on shelter alternatives for finishing cattle.

ALLAN P. RAHN, Ph.D. Iowa State 1973. A quarterly simulation model of the livestock and poultry subsectors for use in outlook and price analysis.

This study develops a systematic procedure for generating intermediate term outlook information for use by decision makers in the five major meat producing industries in the United Utates. These five industries include those producing beef, pork, lamb and mutton, broiler and turkey meat. The procedure utilizes a 47-equation open structural system econometric model which is designed to be operationally manageable and capable of satisfying the "online" temporal requirements of an active outlook information program.

JAVAD M. SADEGHI, Ph.D. Michigan State 1972. An economic analysis of alternative corn irrigation systems in southwest Michigan.

MICHAEL A. SALANT, Ph.D. Stanford 1973. A strategy for operating the upper Pampanga irrigation project: A case study in dynamic optimization involving risk.

An adaptive stochastic model for a dam-reservoir irrigation system is constructed and solved to determine the optimal planted area, throughout the crop year, for irrigated rice production. The model overplants initially to take advantage of possible wet years, but cuts back to avoid crop failure if the year is dry. The adaptive stochastic model is compared, using historical simulations, with various nonadaptive stochastic models and with a deterministic model. Implications for double-cropping are discussed.

HORMOZ SALEH, Ph.D. Cornell 1973. An econometric analysis of the demand for animal protein in Iran.

Cross-sectional data from four household consumption surveys was used to determine the demand for mutton, beef, poultry, and fish in Iran. The effect of income distribution on the demand for animal protein was analyzed by use of a log-normal frequency distribution in conjunction with the Lorenz coefficient. The analysis indicated that the demand for mutton will exceed internal production by approximately 195 thousand tons annually by 1979.

GRANT W. SCHAUMBURG, Ph.D. Harvard 1973. The Paretian model: An application to the Syracuse air pollution problem.

This thesis tests the Paretian model through an application to the Syracuse, New York air pollution control problem. The Paretian model is a mathematical method of generating Pareto optimal outcomes of a bargaining-oriented public decision process. It can be used either as a descriptive and predictive tool, or as an alternative to conventional benefit-cost analysis. The test results indicate that the Paretian model is a potentially valuable tool for applied economics and systems analysis.

DAVID L. SCOTT, Ph.D. Arkansas 1973. The economics of environmental pollution: The case of the electric power industry.

An investigation into the types of residuals, methods of abatement, and costs of control of pollution for the electric power industry. Both fossil-fuel and nuclear plants are examined and control expenses are estimated in terms of costs that consumers can expect to pay in the form of higher bills. The study includes recommendations concerning the industry and environmental control and an overview of economics and pollution.

EUGENE P. SESKIN, Ph.D. Carnegie-Mellon 1973. The social costs of air pollution.

LEONARD A. SHABMAN, Ph.D. Cornell 1973. Decision making in water resource development and the potential of multi-objective planning: The case of the Army Corps of Engineers.

Decision-making processes for four projects are evaluated with a model of political choice processes including factors affecting information transmission and reception. Use of multiple objective optimization models for project formulation and evaluation is then evaluated. Changes necessary to the success of multiple objective evaluation are in regional organization, personnel, hierarchy and regulations, data management systems, criticism and review, public participation, cost sharing plus authorization and appropriation policy.

SURJIT S. SIDHU, Ph.D. Minnesota 1972. Economics of change in wheat production in Punjab (India).

The study focuses on an evaluation of the change in production and cost relationships of wheat crop resulting from the introduction of Mexican varieties of wheat in the Indian Punjab during the latter half of 1960's, and compares economic efficiency of small versus large wheat farms and tractor-operated versus nontractor operated ones. The neoclassical theory of production and profit functions is the basic tool for the analyses.

GEORGE W. SILVERTHORNE, Ph.D. California 1973. Optimal production from a seaweed resource.

DUNSTAN S. C. SPENCER, Ph.D. Illinois (Urbana) 1973. The efficient use of resources in the production of rice in Sierra Leone: A linear programming study.

PHILIPPUS H. SPIES, Ph.D. Iowa State 1973. Management of water quality through selected institutions and instruments.

Water quality management is discussed in terms of planning, organization-coordination and control. Arguments of planning are discussed by means of a model as applied to the Upper Skunk River Basin in central Iowa. Instruments for water quality control are compared. It is concluded that charges for point pollutants and treatment standards for diffussed sources are the most efficient. A general P.T.O. framework for organizing and coordinating water quality management is proposed. The basis of this proposition is free group formation and a hierarchy of constitutions.

THOMAS H. STAFFORD, Ph.D. Cornell 1973. Methods and costs of distributing beef to the food-service industry.

An exploratory cost study of the more common channels of distributing beef to the institutional market. The more common practices and requirements of food-service establishments, beef handlers and packers are identified and examined in separate chapters. A comparison of nine common systems together with a conceptualized cost model serves to identify possible areas of gaining distribution efficiencies. The study concludes with suggested areas for further research in distributing beef to the food-service industry.

ROSCOE E. STOUT, Ph.D. Iowa State 1973. Land as a competitive asset and capital proxy.

An analysis of the degree of substitution between farm real estate and financial assets and the implications of a strong substitution relationship toward the strength of monetary policy. The demand for farm real estate is estimated with time-series data with statistical tests of significance and constancy of regression coefficients over time. A three-asset general equilibrium model is tested using land as a proxy for capital.

SAMUEL M. STRONG, Ph.D. Illinois (Urbana) 1973. An optimal control approach to the off-farm labor migration problem.

MOHAMMED I. SULIEMAN, Ph.D. Missouri 1972. Economic implications of the withdrawal of diethylstilbestrol from cattle feeds.

CHARLES R. TAYLOR, Ph.D. Missouri 1972. Dynamic evaluation of pest control strategies.

GORDON T. C. TAYLOR, Ph.D. California (Berkeley) 1973. An economic analysis of governmental air quality controls.

LLOYD D. TEIGEN, Ph.D. Michigan State 1973. Costs, loss, and forecasting error: An evaluation of models for beef prices.

REMIGIO D. TORRES, Ph.D. Minnesota 1972. Potential benefits and pricing of irrigation water: A case study of the Santa Cruz system.

An economic model was developed to determine the net contribituion of irrigation water to rice output under different levels of rainfall availabilities. The computed value amounted to about P1,275 per hectare during dry season cropping and P934 for the wet season. Shifting the cropping period by one month earlier than traditional starting dates would result in irrigating an additional 1,100 hectares during the dry season. The monetary net return associated with this shift was equivalent to P1.36 over the entire irrigation system of 4,000 hectares. The shift calls for a degree of water management higher than what now prevails.

TIMOTHY F. TREGARTHEN, Ph.D. California (Davis) 1972. Air pollution: An economic analysis.

ALAN H TSAO, Ph.D. Iowa State 1972. Spatial equilibrium and regional comparative advantages in American agriculture.

The linear programming model (2,100 x 8,056) of 15 consuming regions and 138 production areas includes all major crops and livestock products. Regional comparative advantages were identified in the components of shadow prices—production cost, transportation cost, and resource opportunity cost. The two solutions were analyzed on different levels of land and labor constraints (commercial farms vs. all farms): Production shifts to areas with strong comparative advantages. Small farms contribute resources to production in most areas.

STEPHEN E. VAHOVICH, Ph.D. Temple 1973. A framework for evaluating the effect of environmental management policies.

BELLUR V. VANKATAKRISHNA, Ph.D. Kansas State 1972. An economic analysis of processing whole milk into butter and by-products with the continuous churn in Kansas.

PURUSHOTAM L. WAHI, Ph.D. Illinois (Urbana) 1972. An econometric analysis and temporal price equilibrium of U.S. soybean oil.

WILLIAM A. WARD, Ph.D. Michigan State 1972. Native models for estimation of indirect employment and profit gains resulting from public water project investment.

KARL E. WEDEMEYER, Ph.D. Southern California 1972. Interstate natural gas supply and intrastate market.

This dissertation analyzes the interrelationship between jurisdictional and nonjurisdictional markets for natural gas. Interstate contract prices were constrained by Federal Power Commission price ceilings from 1966 thru 1969. Field markets for gas were basically competitive over this period. Excess demand appeared in the interstate market during the early 1970's. The study concludes that the interaction of jurisdictional and nonjurisdictional price and volume behavior affects the outcome of regulatory decisions.

Manpower, Labor, and Population; Including Trade Unions and Collective Bargaining

JOHN P. ADAMS, JR., Ph.D. Claremont School 1973. Training the hard-core unemployed: An evaluation of a union sponsored MDTA program.

This dissertation involves an economic evaluation of the Transportation Opportunity Program—an MDTA demonstration project designed to improve labor market prospects of the severely disadvantaged minority unemployed in the Los Angeles area, with the Southern California Teamsters providing major support for training administration and job placement. Project evaluation entailed estimation of program pay-out for the individual trainee, and examination of trade union support as a determinant of program success.

ROY D. ADAMS, Ph.D. Illinois (Urbana) 1972. A two-sector model of doctorates in higher education.

ROBERT F. ALLISON, Ph.D. Michigan 1972. The role of the nurses' aide.

GOSPEL S. ANGAYE, Ph.D. Pittsburgh 1972. Population and economic development in Nigeria.

This dissertation defines the major demographic and economic features of Nigeria and examines very critically the implications of population size, structure, and rate of growth for Nigeria's economic development. The results of a computerized economic-demographic model show that reduced fertility is favorable to economic and social development in Nigeria.

RICHARD B. ANKER, Ph.D. Michigan 1973. Socioeconomic determinants of reproductive behavior in households of rural Gujarat, India.

MICHAEL W. BABCOCK, Ph.D. Illinois (Urbana) 1972. Employment implications of alternate federal spending priorities for the Illinois economy.

CLIFFORD H. BAKER, Ph.D. North Carolina State 1972. The lagged effect of minimum wages on teenage unemployment.

The lagged effects of minimum wage legislation on teenage unemployment from 1948 through 1968 are investigated by use of a distributed lag econometric model and by use of a partial adjustment model. The distributed lag model is critically examined and shown to possess certain deficiencies in explanatory power. The partial adjustment model is then developed and shown to overcome some weaknesses of the distributed lag model and to improve the understanding of the adjustment process.

THOMAS A. BAROCCI, Ph.D. Wisconsin (Madison) 1972. The drop-out and the Wisconsin apprenticeship program: A descriptive and econometric analysis.

This study examines the various determinants of an individual's success or failure in an apprenticeship program, and investigates whether or not cancellation of the indenture really represents "failure." I employ econometric models to predict the likelihood of success in an apprenticeship, by taking full account of the apprentice's personal and occupational characteristics, as well as opinion indices and labor market indicators corresponding to the time of the apprenticeship. Pre-apprenticeship experience and education and post-apprenticeship employment and income data are also used in the analysis.

ROBERT D. BROGAN, Ph.D. North Carolina State 1972. The determinants of interindustry wage variation.

Average hourly earnings variation by occupation across industries is examined. Three hypotheses are tested in a model based on labor-leisure tradeoffs and labor productivity. The ability to pay hypothesis is in general rejected. An establishment size hypothesis based on work atmosphere disutility is accepted for semiskilled workers. The evidence is generally consistent with a quality difference hypothesis. Capital intensity has predictable effects upon wages by skill class across industries.

ELVIN O. CAMPBELL, Ph.D. Oklahoma State 1972. An evaluation of the performance of workers with impairments at the Oklahoma City air material area.

An annual performance appraisal standard score is used as a surrogate for labor productivity. Impaired workers are compared to nonimpaired control groups by annual wage ranges ($1,000 intervals) and also by occupational groups. The establishment where the study was conducted is a mass employer of the impaired, thereby permitting comparison by impairments which could be treated statistically. The t-test was used to determine the significance of the results.

THOMAS M. CARROLL, Ph.D. Syracuse 1973. Educa-

tion and the distribution of income and wealth among individuals.

The thesis presents a theoretical model with two labor groups: one possessing a single skill, the other also a second skill acquired through education. A homogeneous output can be produced by a production function which uses natural labor skill or another requiring "educated labor." After relating education and labor heterogeneity through the asymptotic distribution of income, attention is given to distributional effects of private and public educational finance.

CRAIG G. COELEN, Ph.D. Syracuse 1973. Properties of the adjustment equation in a model of the demand for workers.

A number of studies, using optimal control techniques, have been made of the properties of adjustment equations in investment. These studies can be extended by analogy to adjustment equations in models of employment behavior. A simple optimal control model of employment behavior is developed in this paper. Explicit adjustment equations are derived for the case in which labor requirements are expected to fluctuate seasonally and secularly over the planning horizon.

C. MARIO M. CORTES, Ph.D. Washington (St. Louis) 1973. Technological absorption and unemployment: A comparative study.

ROBERT W. DACUS, D.B.A. George Washington 1971. Minimizing unemployment welfare costs through the stimulation of unskilled labor employment.

To define a general wage policy which will permit welfare reform while maintaining economic incentive to work if able, manpower and welfare programs, along with pertinent wage theory, are reviewed and a general wage policy is defined providing for the following: a. guaranteed annual income; b. all-inclusive minimum wage; and c. wage subsidies to all workers at minimum wage level. A mathematical model is derived to optimize parameters of above program. *BLS* and census data are used to estimate costs of the program.

JULIE S. DAVANZO, Ph.D. California (Los Angeles) 1972. A family choice model of U.S. interregional migration based on the human capital approach.

This model of migration as an investment in human capital has the following features: financing costs play an important role in determining the optimal investment in migration; for married persons the decision-making unit is the family rather than the individual; in the empirical analysis a wider range of demographic groups, including women and nonwhites, is considered; and some new explanatory variables, e.g., an income-distance interaction term and military indices, are considered (and are statistically significant) in the econometric model.

GORDON W. DAVIES, Ph.D. Michigan 1972. An economic-demographic simulation model designed to test the effects of changes in the rate and skill composition of net immigration on the Canadian economy from 1952-1968.

DOUGLAS N. DAY, JR., Ph.D. California (Berkeley) 1972. Mediation: A multi-stage learning model.

CARL L. DYER, Ph.D. North Carolina State 1972. College and university student migration.

The objective was to study the migration of college students in the United States using 1963 cross-section data by state. Three alternative ways of defining migrants led to three models. The results indicated that tuition and fees, per capita income, and number of colleges in the home state were important determinants of migration across all models. School quality and student ability within the home states were significant variables in some of the models.

KARL A. EGGE, Ph.D. Ohio State 1973. White-black differences in hours of work supplied among men 45 to 59 years of age.

The purpose of this research was to determine if and why the labor supply of black men differed from that of their white counterparts. The data were the 5,000 men 45 to 59 years of age in the first survey (1966) conducted by the U.S. Bureau of Census for the Center for Human Resource Research at the Ohio State University.

THRÁINN EGGERTSSON, Ph.D. Ohio State 1972. Economic aspects of higher education taken under the World War II GI Bill of Rights.

In the study human capital theory is used to evaluate the U.S. government's massive investment in higher education under the World War II GI Bill. The chief data source is a 1967 Census Bureau survey of veterans. Findings include estimates of the impact of the GI Bill on the nation's stock of human capital; regression analysis is used to explain the variance in earnings of veterans; and rates of return to college for veterans are found.

NAZAR S. EL-ISSA, Ph.D. Maryland 1972. Urban unemployment and labor market structure in the process of industrialization.

NADIA H. ELSHEIKH, Ph.D. Oklahoma State 1972. Graduate education for economic development: An evaluation of patterns and trends of specialization of Latin American students in the United States.

The major objective of the study is to discuss and evaluate the patterns and trends of specialization of graduate students from selected Latin American countries (Argentina, Brazil, Chile, Colombia, Mexico, Peru, and Venezuela) in *U.S.* universities in terms of their relevance to their home countries' development needs and employment opportunities. The second objective is to compare and contrast the contribution of each of the graduate students' sponsor groups to the home countries' economic development through their influence on students' specialization.

ALLAN FELS, Ph.D. Duke 1972. The British Prices and Incomes Board: The NBPI as an instrument of incomes policy administration.

PETER FEUILLE, Ph.D. California (Berkeley) 1973. Police union power.

GARY S. FIELDS, Ph.D. Michigan 1972. A theory of education and labor markets in less developed countries.

BOB G. FIGGINS, Ph.D. Arkansas 1973. Intergenerational occupational socioeconomic status mobility of selected 1965 Missouri high school graduates.

Workers are divided into groups according to sex, education level, and person reported by the worker as being the most help in the choice of an occupation. Analysis leads to the following conclusions: a difference in status mobility exists that is attributable to the person reported by the worker as the major influence in occupational choice; workers reporting teachers/counselors as major influencers have a higher percentage of upward mobility than other workers.

GADSDEN C. FRAMPTON, JR., Ph.D. South Carolina 1973. Industry wage discrimination: Additional evidence.

This study, based on the 1966 Survey of Economic Opportunity, attempts to measure the nonwhite wage differential of employees within thirteen industry groups.

ROBERT H. FRANK, Ph.D. California (Berkeley) 1972. The supply of labor.

WILLIAM S. FRANKLIN, Ph.D. Texas (Austin) 1972. An analysis of traditional routes of entry into selected construction unions.

This study of bricklayers, carpenters, electricians, ironworkers, plumbers and steamfitters, and sheet metal workers in Atlanta, New York City, and Austin, has shown that admissions policies resemble each other among locals of the same international union, but vary widely intracity among locals of different internationals. Minority groups seeking access to the trades are at a disadvantage compared with whites, because whites are more likely to know people in the unions and to know how to use the traditional routes of entry. Minorities need to undersand how the traditional routes function in order to increase their participation in the construction industry, but there is also a need to develop new routes to expedite minority employment.

DONALD E. FREY, Ph.D. Princeton 1972. Wage and employment effects of collective bargaining in public schools in New Jersey.

A static equilibrium model of the wage and staffing decisions of school districts was developed. Implications of the model were used to specify wage and employment equations. Wage equations estimated from data from 300 districts in New Jersey showed that recent collective bargaining has had a negligible impact on teachers' wages and employment. The only important determinant of teachers' wages is other wages in the labor market where district is located.

CONRAD F. FRITSCH, Ph.D. Cornell 1973. An evaluation of the economic effects on New York agriculture from the extension of unemployment insurance to the agricultural sector.

This study reviews the development of unemployment insurance and the exclusion of the agricultural sector. It assesses the likely cost of unemployment insurance for agricultural workers, to the State Trust Fund, to dairy farmers and to fruit farmers in New York. Dairy and fruit were selected to represent types of farms using large and small amounts of seasonal labor. The agricultural sector estimated cost rate was similar to the statewide cost for presently covered nonagricultural employers.

MILES E. GALVIN, Ph.D. Wisconsin (Madison) 1972. Collective bargaining in the public sector in Puerto Rico.

The first part of the study provides a cursory description of the long-run dynamics of Puerto Rican society. The second part analyzes the legal factors in support of unilateralism, discusses the factors contributing to employee acceptance of unilateralism, describes how the transition to bilateralism is being accomplished, and analyzes the process of collective bargaining in government employment. The conclusion is that resistance by the Puerto Rican government to extension of industrial democracy to its own employees is resulting in social disruption which appears to outweigh the temporary developmental advantages accruing from imposed austerity.

ROBERT J. GASTON, Ph.D. California (Los Angeles) 1972. A human capital model of age at first marriage in the United States: An econometric study.

ROBERT W. GLOVER, Ph.D. Texas (Austin) 1972. Development and upgrading minority contractors: The Atlanta and Houston experiences.

This paper is an empirical study of minority business enterprise in construction. The analysis is based on a survey of black contractors in Atlanta and black and Mexican-American contractors in Houston. This study outlines the characteristics of the minority contractors, describes their future plans and aspirations, details the problems they face in improving or expanding their businesses. Three approaches to upgrading the minority contractor are examined: government assistance; joint venturing and mergers; and the minority association. Recommendations are made for future public policy.

CLAUDIA D. GOLDIN, Ph.D. Chicago 1972. The economics of urban slavery: 1820 to 1860.

Urban slavery in the antebellum South declined during the last decade before the Civil War. This led many to infer that urbanization and slavery were in-

compatible and that the South did not urbanize because of slavery. I find that the demand for urban slaves was strong during the period 1820 to 1860 and that slaves were pulled out of the cities by rising profits in agriculture. The remainder of the dissertation treats other interesting aspects of urban slavery.

CLARENCE L. HAM, Ph.D. Southern California 1972. The influence of other peoples' schooling on an individual's income.

A theoretical basis was developed for the expectation that the schooling of other people would affect an individual's income. This theoretical construct showed that the schooling of community members should affect an individual's income through market responses and that schooling of the cooperating work group should have effects both through market responses and through spillovers. Multiple regression analysis on the data of the one-in-one-thousand sample of the 1960 Census of Population confirmed these expectations. The effect was relatively more important for minority group members, *certeris paribus*.

GERALD HAREL, Ph.D. Temple 1973. Wage development in Israel 1948–55.

PHILIP J. HELD, Ph.D. California (Berkeley) 1972. The migration of the 1955–1965 graduates of American medical schools.

SYDNEY H. HIGH, Ph.D. Duke 1972. Changes in occupational wage differentials in the United States: 1940–1960.

STEFAN N. HOFFER, Ph.D. Illinois (Urbana) 1972. The determinants of women's earnings.

LAWRENCE R. HUDSON, Ph.D. Syracuse 1973. Measurement of human capital in the 1959 U.S. population.

This dissertation presents and compares two methods to value the stock of human capital. One method treats the earnings of the population as if they were returns on a capital instrument, substitutes cross-section data on earnings by age for time-series data, and capitalizes the earnings at some rate of discount. This method has been in use for almost three hundred years in various forms (Sir William Petty, Burton Weisbrod). The other method is to count up the costs expended and opportunities foregone in activities which create the human capital stock.

MICHAEL D. HURD, Ph.D. California (Berkeley) 1972. Changes in labor force participation.

DAVID G. JAQUES, Ph.D. Claremont School 1973. The academic market place for Ph.D. economists.

This thesis analyzes supply and demand flows for academic Ph.D.'s during the 1960's and early 1970's, particularly as they relate to economics. It discusses the trend toward "oversupply." The forecasting methodology is evaluated focusing on the crucial role of market parameter assumptions. Inferences, from survey data, show changes in market perspectives by formulators of the market. Differences in perspectives and reactions are shown for sub-samples such as academic ranking of school.

PAUL E. JURSA, Ph.D. Texas (Austin) 1972. A case study of the impact of out-migration and return on a southern Italian commune (and province).

After stating that one possible solution to the problems of depressed areas is the out-migration and return of a portion of the area's excess population, I construct a model of the impact of out-migration and return on depressed areas. After illustrating the push-pull factors which stimulate out-migration and characteristics of southern Italian out-migration and return, the experiences of a southern Italian commune and province are examined.

WILLIAM J. KAHLEY, Ph.D. Pennsylvania State 1972. Family planning in Colombia: An analysis of cost and achievement.

RICHARD L. KALUZNY, Ph.D. Wisconsin (Madison) 1973. The determinants of household migration: The poor and the nonpoor.

The study examines the migration behavior of households during 1968–69 and 1969–70. A linear probability model is developed using measures for the expected gain from migration, household income, and other demographic characteristics. The initial analysis groups households by role. Subsequent analysis groups households by four poverty levels. Differential responses to migration determinants by poverty level are noted. The model is most successful in explaining the behavior of the poorest households, and least successful when applied to the very rich.

TIM D KANE, Ph.D. Texas (Austin) 1972. The effect of labor market change upon the employment of low-income Mexican Americans: The San Antonio experience.

This study estimates the impact of structural change upon the labor market in San Antonio, Texas between 1963 and 1969. Occupational estimates of the changing pattern of the demand for labor were compared to the occupations in which training with federal subsidy was provided under the Manpower Development and Training Act.

NABIL KHALDI, Ph.D. Southern Methodist 1973. The productive value of education in the U.S. agriculture, 1964 (a cross-section study).

The objective of this dissertation is to measure the productive value of the education of farmers resulting both from a "worker effect" and an "allocative effect." The worker effect is obtained from an estimate of the agriculture production function. The optimum value of inputs is then derived from the estimated production function and the deviations of observed inputs from this optimum are regressed on the schooling of farm manager, research activity, and

farm size with all data on a state average basis. The data of the 1964 U.S. Census of Agriculture are used. Results show both worker and allocative effects are statistically significant and positive.

LARRY L. KISER, Ph.D. Rutgers 1972. Economic efficiency and public vocational rehabilitation policy.

A public finance-human capital theoretical framework suggests the basic rationale for government to intervene in the vocational rehabilitation of disabled persons—the efficient allocation of resources toward rehabilitation and the redistribution of income toward disabled people. This study attempts to assess the concern for efficient allocation of resources by analyzing individual case and state by state data. The results suggest that economic efficiency is an important concern of the public program.

THOMAS A. KOCHAN, Ph.D. Wisconsin (Madison) 1972. Internal conflict and multilateral bargaining in city governments.

A behavioral model of the city government collective bargaining process is developed and tested. The model is based on the assumption that there is a close relationship between the nature of the internal city management decision-making process and the union management collective bargaining process. The major proposition in the model is that the more conflict that occurs among city officials, the more multilateral bargaining that occurs. A nationwide sample of 380 cities that bargain with locals of the International Association of Fire Fighters provided the empirical basis for testing the model.

EDY L. KOGUT, Ph.D. Chicago 1972. The economic analysis of demographic phenomena: A case study for Brazil.

The study deals with certain demographic phenomena peculiar to Brazil, particularly the existence of different types of marriage. Characteristics of the populations in each type of marriage (such as schooling, income, age) are compared. The characteristics of the different types of marriage themselves (such as duration, age at marriage, stability) are also compared. These characteristics suggest the reasons for different choices made by different individuals.

ANDREW I. KOHEN, Ph.D. Ohio State 1973. Determinants of early labor market success among young men: Race, ability, quantity and quality of schooling.

This study tests a series of hypotheses regarding the determinants of early labor market success among out-of-school young men, where the principal measures of "success" are hourly earnings and social status of occupation. The hypotheses are exhibited in a three-equation, recursive model, and focus mainly on intercolor differences in success and the relative importance of schooling and ability. The hypotheses are tested by multiple regression analyses of data from the National Longitudinal Surveys of Labor Market Experience.

PAULINE W. KOPECKY, Ph.D. Houston 1972. The geographic mobility of American economists.

This study is a longitudinal analysis of the migration patterns of a matched sample of 5,412 economists. Two experiments, utilizing regression analysis, investigate the mover-stayer status, and the salary differential from 1964 to 1968. The third experiment uses a gross flows of labor model to determine the direction and magnitude of migration of these economists. Empirical results are then analyzed in relation to the theoretical hypotheses set forth in the geographic mobility model for professional workers. The findings show little difference in the eight observed characteristics of the migrants and nonmigrants; age was of minor importance; and *ceteris paribus* movers receive higher salaries than stayers.

ROGER LAMM, Ph.D. California (Berkeley) 1973. Black union leaders at the local level.

LEE A. LILLARD III, Ph.D. North Carolina State 1973. An explicit solution to the human capital life cycle of earnings model and its application to earnings distributions.

A solution to a life cycle model was used to guide study of earnings and distribution. An earnings function was estimated using IRSS-Eckland data on high school sophomores in 1955 who took achievement tests and responded to a 1970 survey. The function is used to derive earnings distributions. These distributions display the usual skewness. Specific results included: mean and variance of earnings increase with an increase in age; and an increase in schooling increases earnings dispersion.

DARRYL W. LOWRY, Ph.D. California (Berkeley) 1972. The selective migration bias in labor force participation rates.

ROBERT J. MCINTYRE, Ph.D. North Carolina (Chapel Hill) 1972. Nature and causes of the fertility decline in eastern Europe: Structural change and abortion reform from 1955 to 1968.

JANICE F. MADDEN, Ph.D. Duke 1972. The economics of sex discrimination.

JAMES N. MANGUM, Ph.D. Oklahoma State 1972. An empirical analysis of the relation between personal attributes and the duration of unemployment.

The purpose of this dissertation is to test hypotheses regarding the relation between selected personal characteristics and the duration of unemployment. Special emphasis is given to identifying those factors that are associated with relatively longer periods of unemployment. The sample used in the study is composed of several thousand unemployment insurance recipients who received benefits through a single Brooklyn, New York insurance office between August 28, 1967 and January 3, 1969.

WILLIAM L. MARR, Ph.D. Western Ontario 1973. The

economic impact of Canadian inward and outward migration and their determinants, 1950–1967.

This study examines the economic determinants of inward and outward migration for Canada, 1950–1967. It develops the hypothesis that income and labor market conditions in sending, receiving, and competing countries determine migration flows. I find that Canadian labor market conditions and income prospects influence inward, but not outward migration; that these influences in the sending country are unimportant, and that income prospects in competing countries affects inward migration.

EARL A. MOLANDER, Ph.D. California (Berkeley) 1972. The ideology of aerospace program managers.

JURGEN MULERT, Ph.D. Colorado State 1972. Labor-management conflict resolution.

The literature contains numerous analyses of particular elements of distribution, wage, and bargaining theory conceived as expressions of labor-management conflict. This thesis shows the fruitfulness of this approach in a systematic and comprehensive manner by applying it to the major schools and developments in the history of economic thought.

MAZIN M. NASHIF, Ph.D. Nebraska 1972. An interpretive analysis of labor productivity in selected Nebraska manufacturing industries, 1958–1967.

The study attempted to determine dominant causes of labor productivity movements in sixteen Nebraska manufacturing industries. Broad categories of causes of advances in labor productivity were noted and causal hypotheses were constructed. Cross-sectional correlation and regression analysis were employed to test these hypotheses. The analysis confirmed the hypotheses that productivity movements were attributed mainly to highly labor-saving technical advances and to factor substitution.

ROBERT E. OTLEWSKI, Ph.D. Indiana 1972. Economic and psychological analysis of labor mobility: A study of the Indiana secondary school teacher.

This study focuses on the mobility experience of male secondary teachers in the public schools of Indiana and is restricted to nonoccupational geographic job changes of experienced teachers. Its major objective was to examine and integrate into a model of labor market behavior those noneconomic parameters which affect the incidence and pattern of mobility. The mobility experience of teachers also offers the opportunity to examine the allocative mechanism of a specific labor market.

JOHN PAPADAKIS, Ph.D. Clark 1972. On a theory of the staffing ratio: The relation of clerical staff to blue-collar workers in Greek manufacturing.

The concept of an "Operative Staff Function" is proposed as a way to determine the extent to which the composition of labor force affects the efficiency of industrial firms. The model is applied to fourteen Greek manufacturing industries. The way in which the staffing ratio affects productivity differs between industries. Appropriate adjustments in the composition of labor force may increase overall efficiency in a given economic environment without further increases in the use of resources.

ANTHONY C. PETTO, Ph.D. Missouri 1972. An economic analysis of migration for the Ozarks.

This study examines factors influencing out- and in-migration into the Ozarks region during the 1950 and 1960 decades and also the impact of net migration on the growth of per capita income for Ozark counties. An important finding was that out-migration was primarily determined by economic conditions within the Ozarks region and by noneconomic factors at destination areas. Single equation and simultaneous equation models were utilized.

R. H. M. PLAIN, Ph.D. Alberta 1972. Interregional and interindustry differences in the elasticity of factor substitution and the efficiency of labor in Canada.

RALPH A. PURVES, Ph.D. California (Berkeley) 1972. Interstate migration, demographic trends and regional employment growth: An analysis of recent migration to California.

GARY C. RAFFAELE, D.B.A. Harvard 1973. Background and early history of the international organization of masters, mates, and pilots (AFL-CIO).

This study partially fills a void in maritime history by concentrating on the oldest union of deck officers. It is the result of the author's unlimited access to the organization's historical records. A general background of the traditions and legal environment is given. Emphasis is placed on providing a definitive history of the organization from 1880 to 1930. It was found that general theories of the American labor movement were not completely applicable to this organization.

ALI REZA, Ph.D. California (Berkeley) 1973. A study of employment functions and changes in their structure.

WALTER E. RICE, Ph.D. Claremont School 1973. Employment, unemployment and labor force participation of the rural resident poor—a case study of southern San Luis Obispo County, California.

The basic hypothesis tested is that causes of poverty—limited educational opportunity, prejudice, ill health and a lack of family and labor mobility—are the same for the rural and urban poor. By use of a randomly administered questionnaire amongst rural resident poverty population it was determined that for the South County Planning District of San Luis Obispo County, California, the poverty problem was caused by an excessive supply of low skilled workers who maintained residence because of nonjob considerations in the area.

STEVEN C SALOP, Ph.D. Yale 1972. Wage differentials and unemployment.

MICHAEL J. SATTINGER, Ph.D. Carnegie-Mellon 1973. Two-part tariffs, human capital and nursing.

PHILIP M. SCHERER, Ph.D. Missouri 1972. Rural southern residents and the incentives to work.

This dissertation analyzes the effect of the welfare institution on the incentive to work and whether economic dependency is being generated by this institution. By means of a household survey, the rural heads of households are queried concerning eight values which could affect attitudes toward work. Responses are analyzed by three main categories of respondents—the welfare recipient, the nonwelfare poor, and the nonwelfare nonpoor—and selected demographic characteristics.

RICHARD E. SCHOENBERGER, Ph.D. Clark 1973. The determination of Wisconsin teacher salaries under conditions of bilateral monopoly: Some econometric evidence.

In recent years, two important changes have taken place in the Wisconsin teacher market. School districts have become substantially fewer while school teachers have begun bargaining collectively. Three changes have restricted competition from both the buyer and seller sides of the market; therefore, monopoly and monopsony are important factors in the determination of Wisconsin teacher salaries. This conclusion is based on the empirical evidence resulting from the construction and estimation of a bilateral monopoly model.

ELI SEGEV, D.B.A. Harvard 1973. Environmental analysis and decision making: A case study in the manpower area.

A study of the use of environmental analysis in the decision-making process (related to manpower), based on field research done in Israel. It proposes a dynamic approach to manpower analysis. This study demonstrates the need to change the perception of manpower on the part of environmental analyzers and decision makers from a semistatic manpower "pool" to a concept which recognizes that manpower is a subsystem of the larger environment and that it is dynamic in nature.

WILLIAM J. SEROW, Ph.D. Duke 1972. Economic and demographic implications of a stationary population: A case study of the United States.

RICHARD L. SHORTLIDGE, JR., Ph.D. Cornell 1973. The employment and earnings of agricultural graduates in India: A benefit cost case study of G. B. Pant College of Agriculture and Technology.

A cost-benefit analysis of India's initial agriculture university. Benefits are based on data from a questionnaire mailed to all graduates, and costs are based on official university student and budget accounts. Earnings profiles were computed using single equation multiple regression techniques. The study shows high rate of social and private returns; low unemployment rates; a particularly effective program of degree training of village level workers; high return to private and quasi-public employment relative to state government employment.

JAMES P. SMITH, Ph.D. Chicago 1972. The life cycle allocation of time in a family context.

A model is developed to explain the timing of market work of family members. The model shows that hours of work of each family member is determined by variations in price of time of both members, the rate of interest and time preference, and changes in the technology of household production in the course of the aging process. The model is tested using 1967 *SEO* data.

LARRY J. SMITH, Ph.D. Chicago 1973. Black-white reproductive behavior: An economic interpretation.

Utilizes 1967 Survey of Economic Opportunity data to apply the New Economics of the Family" (*NEF*) to analysis of Children Ever Born (*CEB*) by race and census region. The perspective provided by the *NEF* does not "explain" black-white differences in *CEB*, but interesting differences between the races in the relationships between education, wife's wage, income and family wealth and *CEB* are identified and documented by replication across census region.

RICHARD B. SMITH, Ph.D. Iowa State 1972. Pay packages and employment conditions of farm employees in eastern Iowa.

Study focused upon possible reasons for problems of hiring full-time farm labor. Labor management attitudes of farm employers were found to be significantly related to employee turnover and tenure. Employee status as reflected in titles was usually unattractive. Pay package design lacked competitiveness as some common nonfarm benefits were desired more often than received. Employee pay level was below average manufacturing wages. However, pay level was not found significantly related to employee turnover and tenure.

MICHAEL W. STEBBINS, Ph.D. California (Berkeley) 1972. The social science Ph.D. in corporate industry.

EDWARD I. STEINBERG, Ph.D. New York 1973. Upward mobility of low-income workers.

Using data from the continuous Work History File of the Social Security Administration, this study analyzes the firm- and industry-attachment patterns, and the upward mobility patterns, of various demographic groups in the low-income work force between 1965 and 1970. The study focuses, too, on the effects of such variables as firm size and industry growth on worker attachment and advancement, and on upward mobility in three industries in New York City—apparel, banking, and general merchandise stores.

STANLEY P. STEPHENSON, Ph.D. Indiana 1973. Economics of job search: A biracial analysis of job search behavior of urban male youth ages 18–22.

A basic premise of this study is that lack of job search information to youth entering the labor market, especially black youth, increases youth unemployment, and contributes to black/white differences in earnings, unemployment, and mobility. To examine this premise this study developed a simultaneous equation model of job search. The model contained four equations and twenty-four variables, and used survey data of 150 white and 150 black, young, urban, unemployed men in Indianapolis during November, 1971.

GELVIN L. STEVENSON, Ph.D. Washington (St. Louis) 1973. Determinants of the occupational employment of black and white male teenagers.

A theory of heterogeneous labor inputs takes account of (mis)perceptions and subjective and objective aspects of labor supply and demand. This provides theoretical underpinnings for labor market queue analysis. Regressions (1960 Census data) test for determinants of aggregate and occupational shares (and differences in shares) of black and white males 18–19 years old. Competition between these two groups occurs only in service and labor. Structural factors and racial attitudes are important determinants of occupational shares.

VICTOR C. TABBUSH, Ph.D. California (Los Angeles) 1973. Manufacturers' assistance in the provision of training of employees of customer-firms: Theory and a case study.

This study investigates a phenomenon previously ignored in the economic literature: the provision of training by manufacturers to the employees of client-firms. The hypothesis employed to explain this form of investment is that training may raise not only a worker's productivity, but also the value placed on the resources for which the training is developed. The hypothesis is tested by investigating the mechanic-training behavior of manufacturers in the automobile industry.

BEN B. TARWATER, Ph.D. Missouri 1972. The market theory of union growth: A special case of strategic position in the structure of markets.

THOMAS E. TILL, Ph.D. Texas (Austin) 1972. Rural industrialization and southern rural poverty: Patterns of labor demand in southern nonmetropolitan labor markets and their impact on the poor, 1959–1969.

A 100 percent example of southern counties is used to examine the extent of nonfarm job growth in southern nonmetro labor markets in the 1960's. Then a 147 county sample from 10 states is drawn to determine which industries are growing and declining in nonmetro areas and why. Finally, four Mississippi nonmetro counties are chosen for a study of the impact of rural industrialization on the poor.

ALDON A. TURNER, Ph.D. North Carolina State 1972. An empirical investigation of search techniques used by job searchers.

The purpose of this research was to determine which of the two search models presented in the literature, order statistic and sequential search, better explained the search behavior of job seekers. Several tests were developed to differentiate between the two modes of search. The results tend to justify the assumption of sequential search made by several authors studying the Phillips curve.

MANHARLAL D. VYAR, Ph.D. Pittsburgh 1971. Planning for universal primary education under alternative assumptions of fertility patterns and target dates.

DAVID R. WHIPPLE, JR., Ph.D. Kansas 1972. A generalized theory of job search.

A job search model is constructed incorporating previously ignored features such as a non-linear utility function, endogenously changeable job skills, a career time horizon, a preference for the "kind" of work done, and a modified effect of search. Then the optimal vector of decision variables, namely the acceptance wage and the rate of saving and search, is characterized. Using the information generated by the model, comparisons with the existing literature are made and the significance of differences noted.

WALTER E. WILLIAMS, PhD. California (Los Angeles) 1972. The low-income market place.

Exploitation is a frequently offered hypothesis used to account for observed differences in retail activity between high- and low-income areas. This dissertation offers an alternative hypothesis that these differences can be attributed to merchant wealth-maximizing adaptive response to low-income areas. Models are developed to account for differences in household behavior based on income differences; these models form the basis for the analysis of merchant behavior in low-income areas.

DAVID A. WISE, Ph.D. California 1972. Academic achievement and job performance: Earnings and promotions.

Welfare Programs; Consumer Economics; Urban and Regional Economics

DAVID A. AMEEN, Ph.D. Maryland 1973. An analysis of the welfare economic consequences of forced busing.

The purpose of this thesis is to demonstrate that economic analysis can be fruitfully applied to public policy issues even where distributional or political considerations are involved. Consequently, a model was developed for analyzing the welfare economic consequences of a social issue confronting our society today; namely, the issue of forced busing. To demonstrate applicability, the model was used to evaluate the distributional costs and benefits of forced busing for a particular locality—Richmond, Virginia.

KENNETH L. AVIO, Ph.D. Purdue 1972. Two topics in applied microeconomics: Criminal correction policies and statutory interest rate ceilings on consumer loans.

Two policy problems are discussed. Part one presents a model of a prison as a multi-product firm. Optimal correction policies are developed, with rehabilitative training included as a control variable. Part two presents a model of high-risk loan markets. The supply side is represented by a synthesis of Tobin-Markowitz and safety-first portfolio models. Effects of rate ceilings on the price-volume-credit risk structure of loan markets are derived, with application to adoption of the *UCCC*.

SUSAN A. BIES, Ph.D. Northwestern 1972. Commercial banking in metropolitan areas: A study of the Chicago *SMSA*.

The purpose of this study was to develop a model of the distribution of commercial banking activity within metropolitan areas. Population, income, age, nonmanufacturing employment, and size of the bank and its distance to the *CBD* were found to be positively related to demand. Number of competing banks in adjacent subareas, percentage of population nonwhite, and the ratio of employment to population in a subarea were negatively related to demands.

KEITH E. BERNARD, Ph.D. Maryland 1972. A long term projection of the black-white income gap.

This study projects black and white incomes to the year 2000 on the basis of current occupational distributions. Numbers of income projections are made corresponding to various degrees of occupational discrimination. Racial education differences are projected to narrow but still be substantial by the year 2000. Even in a nondiscriminatory situation, a significant gap between white and black incomes is projected to exist by the end of this century.

DOUGLAS E. BOOTH, Ph.D. Washington 1972. On residential location and the provision of local public goods in metropolitan area.

DALE O. CLONINGER, J.B.A. Florida State 1973. An economic analysis of the relationship between law enforcement rates and the rates of certain criminal activities.

This study takes as its objective the analytical and empirical determination of the relationship between law enforcement and the rates of certain criminal activities. It was found that levels of enforcement are positively associated with levels of criminal activities using cross-sectional data. However, the positive association disappears and becomes negative for most crimes studied when changes in enforcement are associated with changes in crime rates. Serious errors of specification "cloud" the empirical results.

CAROLINE P. CLOTFELTER, Ph.D. Alabama (Tuscaloosa) 1973. Giffin Bonds for minority development.

This study incorporates a neglected aspect of minority behavior in a theoretical and empirical approach to autonomous capitalism. The premise is that discrimination increases the group-identification of minority individuals and thus decreases their risk aversion toward group investment. Consequently they represent a potential source of low-cost development capital, or what is called Giffin Bonds. A case study of Israel Bonds and findings on deviant Negro spending patterns lend support to the theory.

WILLIAM E. COBB, Ph.D. Virginia Polytechnic Institute 1973. The economics of shoplifting.

Criminal offenses were said to be a function of probability of apprehension, punishment, a portmanteau variable representing alternative income sources, moral turpitude, etc., and the perception the potential criminal possesses about these variables. Shoplifting was examined first from the standpoint of the retail merchant and the potential shoplifter. It was shown that a profit-maximizing retailer neither should nor does attempt to alter the first three variables included in the offense function, but that the merchant does attempt to affect the thief's perceptions.

STEPHEN COELEN, Ph.D. Syracuse 1973. A model for measuring the benefits of public investment in urban areas: Theory and measurement.

This study attempts to evaluate, both theoretically and empirically, the relationship between water supply and sewerage system investments and land values. The theoretical model is developed in the early sections of the paper. A case study on water and sewerage investments in Nairobi, Kenya is then presented as an initial test of the model.

RICHARD B. COFFMAN, Ph.D. Washington 1972. The economics of fluoridation.

LESTER R. COHEN, Ph.D. California (Los Angeles) 1973. A time allocation model of residential equilibrium.

This research develops a microtheoretical model describing residential behavior, taking into account the influence of time upon utility maximization in the context of all family consumption and labor supply activities. The household is considered to behave as an economy unto itself with interdependent consumption and production sectors, allocating time and capital according to factor and product market efficiency criteria. Included are critiques of the Muth model of residential equilibrium and Becker's theory of time allocation.

KENNETH A. CONEY, Ph.D. Oregon 1973. Consumer information processing: Organization of product comparison information through the cognitive principle of leveling-sharpening.

ANNA M. CRAIG, Ph.D. Wisconsin (Madison). The distributive impact of urban renewal programs: A case study, Hyde Park A and Hyde Park B.

Two Chicago urban renewal clearance and redevelopment projects were analyzed. In these cases, land acquisition and disposition had little redistributive impact. However, prerenewal tenants suffered large net losses relative to their incomes. The major beneficiary was the city government, although because of taxpayer contributions to the federal cost of these and all other federally assisted urban renewal projects, the benefits to the city's citizens was substantially less.

CHARLES B. DANIELS, Ph.D. Stanford 1973. An investigation of the influence of racial segregation on housing prices in the Oakland, California housing market.

Using a regression model, this study examines the relationship between housing prices and residential segregation in a California housing market with data from the 1960. Census and a 1965 land-use survey. White renters paid a segregation premium to live in the white submarket, implying competition rather than collusion in the rental housing market; a unit of housing space was more expensive in the black market while a unit of housing quality cost more in the white submarket; and home owners pay premiums for sites farther from the city central business district where blacks are concentrated. Legislatively imposed quota systems or financial incentives would be more effective than antidiscrimination legislation in reducing residential segregation.

SIDNEY DAVIS, Ph.D. Georgia State 1973. Household consumption of housing service flows in Atlanta, Georgia.

This study examines consumer demand for housing service flows which are represented by the utility derived from a dwelling unit that has the following characteristics: a specific spatial location in the urban community; a specific vector of physical attributes including size and quality, in combination with a quantity of "transportation," enabling the household to participate in activities taking place in locations removed from the place of residence. The housing service flow concept then serves as the basis for a simultaneous equation model.

FRANCIS J. EGAN, Ph.D. Fordham 1973. Air pollution and property values in the Hartford metropolitan region.

This dissertation investigates the relationship between air pollution and residential property values in the greater Hartford metropolitan region. To measure this relationship, a single equation, linear regression model is constructed with property values as the dependent variable and air pollution as one of the independent variables. Census tracts are chosen as the basic units of observation.

JAMES A. M. ELLIOTT, Ph.D. Yale 1972. International comparisons of social welfare: A factor analytic approach.

MICHAEL C. ELLIS, Ph.D. Alabama 1973. An empirical and methodological approach to the relationship between regional income inequality and economic development: A study of the Ozark region in 1970.

A significant and sometimes unforeseen consequence of regional development programs is an increase in the inequality of personal income. In order to ascertain whether a growth-income inequality tradeoff exists for the Ozark region, the area was analyzed cross-sectionally in 1970 to determine: the extent of family income inequality; the major principal component dimensions of development; and the relationship between family income inequality and economic development.

DONALD R. EPLEY, Ph.D. Missouri 1972. A test of base theory using income and employment.

This study developed for the state of Missouri four Laspeyres indices which were incorporated into base theory. Two indices measured income and employment fluctuations for the basic sector and the other two, the nonbasic sector. Index components and weights were determined from the state input-output table. Results revealed that base theory in income terms has no relationship to base theory in employment terms. The roles of the employment indices were reversed.

CHARLES L. FRANCE, Ph.D. West Virginia 1972. A production model of a large city police department.

AL FRANCFORT, Ph.D. Pittsburgh 1972. Driving without a steering wheel—the need for and use of health statistics in economic analysis.

The central objective of this study is to suggest what should be done to obtain the health information needed to achieve a better allocation of resources both *to* and *within* the health sector of the economy. The premise is that the health statistics currently available are inadequate. The study suggests possible improvements in the health information system which should result in additional benefits to health care and increased levels of healthiness in the population.

RICHARD T. FOX, Ph.D. West Virginia 1973. The cost of selected hospital services.

LEE S. FRIEDMAN, Ph.D. Yale 1973. Innovation and diffusion in non-markets: Case studies in criminal justice.

WILLIAM C. GIAUQUE, D.B.A. Harvard 1972. Prevention and treatment of Streptococcal sore throat and rheumatic fever—a decision theoretic approach.

In this research the medical problem of prevention and treatment of Streptococcal sore throat and rheumatic fever was examined and alternative treatment strategies evaluated. Methods of decision theory and multivariate utility theory were used to structure the problem and define an objective function. A procedure to determine parameters of the utility structure was determined and applied to patients, doctors, nurses, and public health officials. Implications of results for clinical practice were discussed.

STEVEN D. GOLD, Ph.D. Michigan 1972. The distribution of recreation services in Detroit 1968–69.

RONALD E. GRIESON, Ph.D. Rochester 1973. Essays in urban economics and public finance.

"The incidence of profits taxes in neoclassical growth model" is a theoretical study of the effect of taxes on the capital-labor ratio, and on the long-run wage, net of taxes. "The economics of property taxes and land values" is a simultaneous model of land values, uses, and gradients; elasticities of supply, price, and densities of structure. "The determinants of juvenile arrests" is an econometric estimate of the effects of social and economic programs.

JOHN A. HANSON, Ph.D. Oregon 1972. A vintage capital model of urban development.

ROBERT G. HEALY, Ph.D. California (Los Angeles) 1972. Agglomeration and footlooseness: The distribution of economic activities among metropolitan areas.

The dissertation deals with the spatial agglomeration of industries, both manufacturing and nonmanufacturing, in 203 metropolitan areas. Factor analysis is used to divide 186 industries into groups which have the property that cities specialized in one industry are likely to also specialize in the others. The role of backward and forward linkages in producing agglomerations is explored and a regression model is used to identify the dominant locational forces for each of several selected industries.

THOMAS L. HEFLIN, Ph.D. Oregon 1973. The prediction of retirement adjustment.

FREDERICK J. HITZHUSEN, Ph.D. Cornell 1972. Some policy implications for improved measurement of local government service output and costs: The case of fire protection.

This study uses a multiple curvilinear regression analysis of fire protection unit costs and output (people and property protected) in 150 Texas and New York localities. Costs include not only public sector current expenditure but estimates of public capital costs, public water supply costs and unpaid volunteer time contributed. A benefit-cost analysis provided estimates of tradeoffs between public fire protection improvements and private fire insurance savings.

JOHN H. HOAG, Ph.D. Kansas 1972. The economics of the health service industry.

The dissertation reports on an economic model of the model to two problems relating to the allocation of resources in this market. The model is an aggregate general equilibrium model involving four sectors—consumers, doctors, hospitals, and insurance. The first application deals with the effect of insurance on the consumer's choices; the second is an investigation of the relationship of the doctor and the hospital.

PETER M. HUTCHINSON, Ph.D. Pittsburgh 1972. Accessability and segregation: Their effects on the employment of the urban poor.

The dissertation uses a linear probability model based on observations of poverty zone residents in the Pittsburgh region to determine how housing segregation and employment accessibility affect the probability of poverty area residents being employed. Results of the empirical tests suggest that improved public transit in the form of a dial-a-bus or a program of providing automobiles for the poor may significantly improve employment prospects for the urban poor. On the other hand, ghetto dispersal policies aimed at eliminating housing segregation does not improve employment possibilities.

LEONARD W. KATZ, D.B.A. George Washington 1972. An evaluation of the state of Connecticut Health and Educational Facilities Authority's impact on the provision of nongovernmental health facilities.

Compared to other reported long-term borrowing, authority tax-exempt revenue bond financing provided hospitals with lower interest rates and cost of capital and longer maturities. Debt principal equalled between 60 and 93 percent of total costs. Hospitals borrowed 36.9 to 89.8 percent of projected debt capacity. Debt servicing might be adversely affected if third party reimbursement declines or the construction period extends materially beyond the expected completion date.

WERNER KIENE, Ph.D. Michigan State 1972. Evaluation of the impact of health care on activity levels of the rural poor.

NEIL KOMESAR, Ph.D. Chicago 1973. Economic analysis of criminal victimization.

PETER J. KUCH, Ph.D. Minnesota 1972. A planning for postsecondary education in west central Minnesota.

This study is concerned with planning the efficient development of postsecondary education in a region. Attention is focused on allocating public resources among competing curricula and institutions, where primary concern is on the pecuniary returns to schooling. The procedure employed uses multi-period linear-programming models which maximize the net social benefits generated by the postsecondary system subject to constraints describing the educational technology available to the region and externally imposed policy decisions.

HAROLD LUFT, Ph.D. Harvard 1973. Poverty and health: An empirical investigation of the economic interactions.

This study provides estimates of the probability of disability based on age, race, sex, education, and predisability income and detailed working conditions. The second half derives estimates of earnings and income losses attributable to disability, as well as the share of poverty caused by poor health. Serious specification errors are demonstrated in labor force equa-

tions without health status. Aggregate and individual data are drawn from the 1967 *SEO* and 1966 Survey of Disabled Adults.

KENNETH E. MCCONNELL, JR., Ph.D. Maryland 1973. Estimating consumer demand functions from cross-section data.

Cross-section estimates of consumer demand functions, though valuable because they are devoid of problems of autocorrelated errors, have been rare because of the absence of price variations in cross-section data. This study attempts to overcome the problem of no price variation. Price elasticities from directly and indirectly additive utility functions are computed by (1) estimating consumers' responses to different wage rates; (2) imposing second-order conditions on structural parameters. Results from elasticities computed from a group of *SMSA's* demonstrate that such elasticities can be useful.

PATRICK A. MCGUIRE, Ph.D. Fordham 1973. A simulation model of the Rochester, New York region.

The object of this study is to develop a simulation model for an urban area. The model will attempt to illustrate the effectiveness of various policy changes on the theoretical city. The objective is to demonstrate that traditional policy recommendations may not be as successful in the long run as expected. The model will be applied to a particular city to illustrate the long-run effects of policy changes.

PETER J. MCMAHON, Ph.D. Southern California 1972. Forecasting selected statewide recreation requirements.

This study develops an approach to forecasting outdoor recreation demand and supply requirements. To provide the background for the approach, the study first analyzes the nature of the recreation market as well as the difficulties, methods used, and relevant factors in making these projections. An empirical model is then formulated using microeconomic constructs to enable forecasts of recreation demand and optimum combinations of selected recreation facilities that must be supplied to meet this demand.

JOHN C. MCPHERSON, Ph.D. Brown 1973. The economics of central place hierarchies: Applications to the great plains experience.

This thesis presents a formal view of the economic structure of urban centers in terms of their spatial, size and frequency distribution. A generalized hierarchy model is formulated as a set of difference equations with explicit spatial and nonspatial parameters, and an explicit solution obtained. Within this framework the parametric differences between Canadian and American great plains economies in 1960–61 are examined, along with the change in spatial parameters in Saskatchewan 1941–61.

JOSEPH P. MAGADDINO, Ph.D. Virginia Polytechnic Institute 1973. An economic analysis of state gun control laws: A statistical study of 1970 data.

The problem of firearm misuse was defined on the basis of an externality argument. To test the efficacy of gun control laws, a cross-sectional study of supply of offenses was conducted. This study attempted to explain the differences in crime rates among states attributable to differences in the level of deterrence, and the gun control laws while accounting for differences in socio-demographic variables.

ADELE P. MASSELL, Ph.D. Stanford 1972. A bimarket analysis of the urban labor market.

This study analyzes change in the location of new plants in suburban areas surrounding central cities, to determine its effect on employment. The economic implications of alternative intraurban spatial configurations of employment are made through a partial-equilibrium, comparative statics model, with two interdependent labor zones. Both central and suburban workers are best off when the spatial distribution of firms parallels that of workers' residences. When racial variables as proxies for labor skill variation are included in the model, results suggest that when firms suburbanize white central and suburban residents tend to benefit but that central nonwhites face lower wages.

MARSHALL H. MEDOFF, Ph.D. California (Berkeley) 1973. A theoretical and positive analysis of discrimination in the market place.

STEPHEN L. MEHAY, Ph.D. California (Los Angeles) 1973. Production functions for crime-deterrent police services.

This study is a theoretical and empirical investigation of the production of crime-deterrent services by urban police departments. A model of police protection output is developed and estimated using cross-section data from southern California police departments.

EMIL MEURER, Ph.D. Nebraska 1972. Crime compensation as an emerging form of social insurance.

This investigation was made to determine the extent and nature of existing state and proposed federal programs designed to alleviate the financial burden of violent crime on the victim. Status of violent crime victim is reviewed and present legislation is evaluated with recommendations for improvements.

JAMES E. MILLER, Ph.D. Texas A&M 1973. Analytical and operational aspects of central facilities location.

The objective of this study is to adapt traditional private facility location theory for central public facility location. The conclusions are as follows: the location theory of Alfred Weber, as a basis for certain operational techniques (e.g., spatial combinatorial programming) may be successfully applied to only small scale problems involving emergency facility location; and Löschian market area theory as a basis for developing an analytical model with the opportunity costs of user travel time introduced as a demand parameter, is a more promising approach.

WILLIAM G. MOSS, Ph.D. California (Berkeley) 1973. The effect of housing segregation on the Negro journey to work.

DAVID H. MUDARRI, Ph.D. Michigan 1972. Mode choice for work trips: A key urban transportation problem.

EDWARD H. MURPHY, Ph.D. Rutgers 1973. The market for disability insurance benefits.

The thesis analyzes the nature of the relationship between the economic environment and the market for disability insurance benefits. A theory of the demand for benefits is derived and tested for several segments of the disabled population.

JAMES A. MURRAY, Ph.D. Oregon 1972. The economics of medical care: a theoretical analysis.

This dissertation presents a positive microeconomic theory of the production and use of medical care. A primary objective is to determine the significance of well-known product and institutional peculiarities. Standard optimization techniques are used to develop theorems from a set of behavioral hypotheses. These theorems describe how pricing and utilization of physicians' and hospital services relate to factor prices, patient incomes and insurance coverage. Results obtained differ importantly from theorems of neoclassical models.

TIM O. OZENNE, Ph.D. California (Los Angeles) 1972. The economics of theft and security choice.

Theft is controlled not only by governmental actions but also by private security choice. This study develops a model of theft in which the rate of theft is determined jointly with optimal security expenditures. The model is then used to explain variation in bank robbery rates across states and over time. The empirical results conform with the theoretical implications and show that earnings net of imprisonment costs have the anticipated effect on robbery behavior.

RICHARD B. PARKER, Ph.D. Boston College 1973. Interregional income distribution, factor migration, and growth.

Analysis, parameter estimation, and computer simulation of dynamic interregional models of migration and income are carried out to explore the equity-efficiency tradeoff. At very low levels of factor migration response to productivity differences, a policy which increases their response will improve interregional equity. A maximum is found, beyond which further increases in the migration-reaction coefficient result in equity loss. Below this maximum, higher growth under an equity constraint requires greater factor migration response.

PAUL U. PAWLIK, Ph.D. Arizona 1972. The effects of health fluctuations on economic activity in Iran.

The question examined is whether or not fluctuations in health, in Iran, affect individual firm's production, individual industry's production, and general economic activity. Fluctuations in health will be compared statistically to fluctuations in individual firm's production, individual industry's production, and general economic activity over a four-year period (on the monthly basis).

NORMAN E. PAYNE, Ph.D. Cornell 1973. An economic evaluation of school lunch systems.

With school lunch programs receiving increasing attention for social and nutritional reasons this research focused upon the basic economic aspects of alternative systems. Wage rates and labor productivity were clearly the most important factors causing variations in the cost of providing a school lunch in seven eastern and southern states. The type of system (unit kitchens or satellite facilities) and size of system were much less important.

ROBERT M. PETERS, Ph.D. Kentucky 1972. A study of teacher aide impact and the economics of aide usage.

JOHN G. PHOTIADES, Ph.D. Illinois (Urbana) 1972. A multivariate analysis of family income distribution of U.S. counties.

WILLIAM POLLAK, Ph.D. Princeton 1972. Public fire protection: Costs, demands, and pricing.

Suppression capability is defined as a quality measure. A priori models drawing on identified response-time/damages relationships are used to analyze the area, population, and quality dimensions of the protection cost function. A firespread model is used to analyze protection demands and the efficiency benefits of pricing protection are examined using geometric models. These show that in a dynamic setting fees may prevent an over shifting of protection burdens to the public sector.

ELIZABETH A. ROISTACHER, Ph.D. Pennsylvania 1972. The distribution of tenant benefits under rent control.

FREDERICK H. RUETER, Ph.D. Carnegie-Mellon 1973. Municipal zoning: An empirical study of its economic and political justification.

DANIEL H. SAKS, Ph.D. Princeton 1973. Economic analysis of an urban public assistance program: Aid to New York City families of dependent children in the sixties.

Using primarily survey data for New York City in 1967, creation of welfare cases was analyzed as a problem of rational choice. The welfare system was found to be a complex negative tax scheme with tax rates well under 100 percent. Plausible income and substitution parameters were estimated for supply-of-recipients, on-welfare-labor-force-particiaption, and welfare-duration equations. Welfare survival rates were estimated and shown to be Markovian.

C. JAMES SAMPLE, Ph.D. American 1972. Patterns of economic change in regions of the continental United States.

JAMES L. SHANAHAN, Ph.D. Wayne State 1972. Spatial isolation and job opportunity for low skill residents in the core of the Detroit SMSA.

TIMOTHY G. SHANAHAN, Ph.D. Purdue 1972. Equity and efficiency of alternative policies and mechanism for optimizing consumer-type levies.

RAM K. SHARMA, Ph.D. Western Ontario 1973. Benefit-cost analysis and public health: A case study of the Tuberculosis Control Program in Ontario, 1948–1966.

This dissertation develops and explains the methodology of applying benefit-cost analysis to public health projects and then applies it to a specific cases: The Tuberculosis Control Program in Ontario from 1948 to 1966. The marginal benefits and costs of the program, as measured in this study, prove that the program was a profitable use of resources even after accounting for the beneficial effects of "wonder drugs" discovered for tuberculosis treatment.

ALLAN K. SHERMAN, Ph.D. Maryland 1972. The effects of the configuration of rural health care systems on elective utilization.

MARK K. SHERWOOD, Ph.D. Purdue 1972. The demand for public and private higher education.

This thesis is a theoretical study of the choice a qualified high school graduate makes among attending a public college, attending a private college and entering the labor force.

GAD SHIFRON, Ph.D. Wisconsin (Madison) 1972. Income conditioned publicly subsidized preschool child care.

The objective of this study is to determine the economic consequences of income conditioned public subsidization of preschool child care services. It is concluded that deducting child care expenses from taxable income can be justified, but income conditioned tuition is hard to justify. It is demonstrated, on equity and efficiency considerations, that day care centers may be an expensive while not the most effective solution for improving the opportunities of mothers and children. The policy implications suggest a reconsideration of child care provisions in H.R. 1 and S-3617 (both are major bills pending in Congress at the present time).

LESTER P. SILVERMAN, Ph.D. Carnegie-Mellon 1973. The determinants of emergency and elective admission to hospitals.

JOHN C. SIMONSON, Ph.D. Wisconsin (Madison) 1972. The determinants of changes in the allocation of hospital service capacity: A study in policy theory.

LAWRENCE T. SMEDLEY, Ph.D. American 1972. The impact of a social security program and a number of other factors on household savings in a developing country: The Mexican case.

MICHAEL W. SMITH, Ph.D. North Carolina State 1972. An economic analysis of the intracity dispersion of criminal activity.

The spatial distribution of, and economic causes of, criminal activity within a metropolitan region was studied by considering the crime flows between subsections of the region. A theoretical model was developed to predict the magnitude of the import, export, and domestic production of property crime relative to a subsection. These quantities were assumed to be dependent on the income of offenders and victims of crime, the distance between subsections and the size of the subsections in terms of the number of potential offenders and victims. The results of a regression analysis indicated that the model explained a significant proportion of property crimes.

DONALD N. STEINNES, Ph.D. Northwestern 1972. An econometric model of residential location.

In this thesis a model is derived which explains the intraurban location of both residents and employment and yet is capable of being estimated from available data. Particular attention is given to the simultaneous nature of the geographic allocation of employment and residences within a metropolitan area, and to the economic-theoretic basis of these allocations. The model is estimated by *OLS* and *TSLS* using data for the Chicago metropolitan area.

STANLEY C. STEVENS, Ph.D. Illinois (Urbana) 1972. The spatial dynamics of the Chicago black community: 1958–69.

MICHAEL SUNSHINE, Ph.D. Michigan State 1973. The abuse of warranty law as a contributing factor to consumer problems.

RAYMOND J. SUPALLA, Ph.D. Michigan State 1972. The potential for increasing earned income of welfare recipients in rural areas: A case study of Allegan County, Michigan.

TATSUO R. TSUCHIGANE, Ph.D. Maryland 1972. Discrimination against women in the United States.

This study is an attempt to measure the components of discrimination against women in the economy and to identify factors affecting this discrimination. Each of the three types of discrimination, namely, income discrimination, occupational discrimination, and participation discrimination, is separated into "justified" and "unjustified" components. By separating discrimination into two components it is possible to gain insights into the extent of justified and unjustified discrimination and the prospects for removing unjustified discrimination in the future.

BRADFORD H. TUCK, Ph.D. Boston 1973. The economic impact of the petroleum sector on the state of Alaska, 1960–1970.

... IL TUDIVER, Ph.D. Michigan 1973. Why aid doesn't help: Organizing for community economic development in Central Appalachia.

MICHAEL A. VIREN, Ph.D. California (Santa Barbara) 1973. An economic analysis of the supply of medical services.

ROBERT A. WEISS, Ph.D. Wisconsin (Madison) 1972. The distribution of benefits of the Appalachian Development Highway System.

SHIRLEY F. WEISS, Ph.D. Duke 1973. New town development in the United States: Experiment in private enterpreneurship.

RAFAEL R. WESTON, Ph.D. Harvard 1972. The quality of housing in the United States, 1929–70.

Housing, like all capital goods, must be treated from the points of view of both stocks and flows, and the relationship between the two is crucial. A Divisia-hedonic model of housing indexes incorporating this distinction is developed, and empirically implemented to investigate the question of housing quality. This analysis is then employed in the creation of a complete, internally consistent, system of accounts for the housing sector of the economy.

ROBERT S. WOODWARD, Ph.D. Washington (St. Louis) 1972. Industrial incentives and regional economic growth: The case of Puerto Rico.

A priori profit impacts for each of the three intraisland industrial incentives were first estimated for typical export manufacturing firms. The empirical significance of governmental construction of standard industrial plants and extra years of tax exemption, which was indicated by profit impact calculations, was verified using promotional employment data in cross-sectional *OLS* regressions.

ANNOUNCEMENT

NOTICE TO ALL GRADUATE DEPARTMENTS

The December 1974 issue of the *Review* will carry the seventy-first list of doctoral dissertations in political economy in American universities and colleges. The list will specify doctoral degrees conferred during the academic year terminating June 1974. This announcement is an invitation to send us information for the preparation of the list. This announcement supercedes and replaces a letter which was sent annually from the managing editor's office.

The *Review* will publish in its December 1974 issue the names of those who will have been awarded the doctoral degree since June 1973, the titles of their dissertations, and, if possible, a brief (75-word) summary of the dissertation.

By June 30, please send us this information on 3×5 cards, conforming to the style shown below, one card for each individual. Please indicate by a classification number in the right-hand corner the field in which the thesis should be classified. The classification system is that used by the *Journal of Economic Literature* and printed in every issue.

JEL Classification No. ______

Name: LAST NAME IN CAPS: First Name, Initial

Institution Granting Degree: ______

Degree Conferred (Ph.D. or D.B.A.) ______ Year ______

Dissertation Title: ______

Summary

(75-word maximum, or first 75 words will be printed)

Summary may be completed on back of this card or on new card which should be stapled to this.

When degrees in economics are awarded under different names, such as Business Administration, Public Administration, or Industrial Relations, candidates in these fields whose training has been *primarily in economics* should be included.

All items and information should be sent to the Assistant Editor, *American Economic Review*, Box Q, Brown University, Providence, Rhode Island 02912.